The Dissertation Committee for Gabriel Antonio Suárez Certifies that this is the approved version of the following dissertation:

Simplified engineering of *Acinetobacter baylyi* ADP1 and evolutionary strategies to genome minimization

Committee:

Jeffrey Barrick, Supervisor

Hal Alper

Nancy Moran

Kyle Miller

Bryan Davies

Simplified engineering of *Acinetobacter baylyi* ADP1 and evolutionary strategies for genome minimization

by

Gabriel Antonio Suárez

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August 2019

Acknowledgements

All throughout my graduate school journey, I've been blessed by the support of several remarkable individuals.

First, I'd like to thank my advisor, for lending me the vision and guidance to pursue with great scientific rigor many challenging, yet fruitful, explorations.

I'm deeply grateful for the advice, discussions and insights to help me make the best of my research project from all participating members of my dissertation committee, namely, Drs. Hal Alper, Kyle Myler, Nancy Moran, Bryan Davies and retired Dr. Richard Meyer.

I would like to thank family and friends; life gave me so much energy from each and every one of you, José Felix, Rubén, Alfredo and tía Marta, in particular.

I would like to thank all my lab colleagues, each in some way helped build the strengths I now carry with great pride.

To my dear son, Antuan, for instinctively knowing how to bring me back home and rebuild my energies, always.

I give thanks to the infinite supply of strength from my mother.

Lastly, Hollie, my blissful refuge, here's some evidence of all the time I owe you.

Simplified engineering of *Acinetobacter baylyi* ADP1 and evolutionary strategies for genome minimization

Gabriel Antonio Suárez, Ph.D.

The University of Texas at Austin, 2019

Supervisor: Jeffrey E. Barrick

Our ability to engineer and domesticate microbes to give them useful properties promises grand rewards in the energy, agriculture, chemical and health industries. Yet, synthetic biologists often struggle to engineer bacterial genomes despite ever-improving genome-scale models of how they function. Often, they are stymied by the sheer complexity of the cell's underlying systems biology and by how these continue to evolve rapidly after they are engineered. Recent advances in genome stabilization and genome simplification promise to overcome these barriers and profoundly extend our understanding of basic molecular biology and cellular life. Both the natural instability of bacterial genomes and their unexplored complexity (e.g., the presence of many genes with unknown functions) underlie major challenges to be reckoned with that often lead synthetic biologists to rely on extensive experimental trial and error.

The construction of cells with minimal genomes to make microbiology more predictable is riddled with difficulties. There are sometimes advantages and sometimes disadvantages for removing more and more genes to simplify a bacterial cell. Similarly, evolution is a process that may both frustrate or enable synthetic biology. It can be slowed down by removing selfish DNA elements from a genome or it can be applied to compensate for suboptimal designs. The work in this thesis explores these interactions

between genome design and evolution. It asserts that rational engineering and simplification principles can lay stronger foundations for engineering microbial cells so that more complex and ambitious designs can be successfully built, but that evolution is also a necessary tool to achieve extreme simplification of a living cell to make it robust enough for research and industrial demands to achieve the potential of synthetic biology.

Our model organism is *Acinetobacter baylyi* ADP1, a highly naturally transformable and metabolically versatile soil bacterium.

Chapter 1 provides an introduction to *A. baylyi* genetic engineering and the current state-of-the-art in bacterial genome stabilizing and streamlining projects. **Chapter 2** describes our rational engineering efforts to reduce *A. baylyi* ADP1 genome instability–mainly by deleting all transposable elements from its genome–and the beneficial phenotypes in the ADP1-ISx strain that resulted from this work. **Chapter 3** describes improved *A. baylyi* genome engineering methods and how they were used in the first stage of a genome streamlining project. We also describe a "Golden Transformation" protocol that speeds up and simplifies the steps needed to make precise edits to the *A. baylyi* genome and also show that the native CRISPR-Cas system is functional and can be reprogrammed using this method. **Chapter 4** describes how we begin to test how compensatory evolution of reduced genomes can open new pathways to more extreme genome minimization by restoring fitness that is lost after deleting many dispensable genes from a genome. **Chapter 5** discusses future directions for making improvements that further stabilize and streamline the *A. baylyi* genome.

Together, the work presented in this dissertation presents concepts, tools, and insights into strategies that were successful and unsuccessful for building a better and simpler *Acinetobacter baylyi* ADP1 genome. These approaches can also be applied to other bacterial species to propel the goals of synthetic biology forward.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Acinetobacter baylyi ADP1 as a platform for biotech applications

A. baylyi ADP1 is a non-pathogenic, strictly aerobic, soil γ-proteobacterium that is notable for its high natural transformability, metabolic versatility, and genetic tractability [1,2]. It has a truly remarkable ability to indiscriminately uptake and incorporate diverse sources of DNA into its genome during normal laboratory growth [3–9]. This ability makes it straightforward to edit its genome sequence [10]. ADP1 also has a well-developed set of resources, including a single-gene knockout collection and genome-scale metabolic models, to guide engineering its capabilities [11–14]. ADP1 offers broad opportunities for metabolic engineering [15], sensor development [9], and asking fundamental questions about competence mechanisms [6,16] and genome evolution [17–22].

Broad metabolic versatility is fueling ever more studies dedicated to metabolic rewiring of *A. baylyi* ADP1 [23]. Deletion of genes for competing metabolic pathways has been used to optimize ADP1 for the production of compounds with applications in the food, cosmetics, and biofuel industries, such as alkanes [24,25], triacylglycerol [26] and wax esters [27 28] . In one case, where metabolic reprogramming shifted the carbon source from succinate to quinate, an engineered ADP1 strain produced dozens of unidentified, and possibly novel, compounds [13]. Similarly, a single-gene knockout that prevents sugar/glucose metabolism in ADP1 makes it an effective biodetoxification tool for the removal of oxygen that would otherwise inhibit anaerobic fermentations [29]. Beyond this

sort of "reductive" metabolic engineering, construction of synthetic pathways in ADP1 has been equivalently fruitful, recently exemplified by work increasing production of fatty aldehydes [30].

Early on, it was shown that ADP1 bioreporters could be designed and employed for detection of a wide range of compounds, for example, salicylates [31–33], alkanes [24,34] and toxic compounds [35,36]. In one of these cases, twin-layer biosensor cells were developed that could detect both degradation and biosynthesis of alkanes [24]. This is possible because ADP1 is responsive to alkanes as signal molecules to direct transcriptional regulatory proteins (e.g., AlkR) [37]. Due to its high competence, ADP1 has also been engineered as a DNA biosensor to detect marker genes from transgenic plants [5].

Its high natural transformability and propensity for evolving gene amplifications in its genome, yield unique opportunities for the directed evolution of useful pathways in *A. baylyi*. Transformation-facilitated PCR mutagenesis in ADP1 can generate useful locus-specific mutation libraries for structure-function studies and improving protein properties [38,39], as has been shown in the case of the transcriptional activator (*pobR*) of the *pobA* gene. Purposeful emplacement and amplification of heterologous genes involved in lignin degradation in the genome have provided opportunities for the evolution of improved enzymes, including the generation of protein chimeras with improved catalytic dynamics [40]. These capabilities enhance the usefulness of ADP1 as a study system for metabolic engineering, genome evolution, and synthetic biology.

Evolutionary challenges for synthetic biology

Synthetic biologists have long sought to create useful genetic circuitry to exploit the metabolic, signaling, and sensing capabilities of cells; their progress has immense potential for products and breakthroughs that improve human lives [41] [42] [43]. One challenge with making these a reality is that many of these devices are evolutionarily short-lived. They often mutate rapidly and these mutations can quickly dominate cell populations. This failure mode constrains attempts to develop better and more complex biological systems. For example, populations that highly express fluorescent proteins mutate this burdensome function within a relatively small number of cell divisions [44]. The natural evolvability of genomes makes it difficult to sustain burdensome and complex genetic circuitry that is non-native to a host cell. Point mutations, deletions, and insertions – in addition to mutations caused by selfish DNA elements such as transposons and prophages – can frustrate synthetic biology efforts [45].

Another stumbling block to synthetic biology is the sheer complexity of the cellular milieu in which an engineered circuit or pathway is deployed. An ideal chassis organism would be one that is thoroughly understood, such that when engineered circuits are added, these encounter no surprise interactions with host functions [46] [47]. Natural organisms have evolved to thrive in many different environments and survive many different stresses. Much of this machinery (e.g., accessory genes, genes involved in non-central metabolism and certain membrane components) may not be necessary in an application in an industrial fermenter (**Fig. 1.1**). They may even waste energy or interact in unpredictable ways with engineered components. These notions, along with the

realization that organisms with severely reduced genomes can be found living stably in natural—highly static—environments (e.g., *M. genitalium* with a < 580 kb genome) (reviewed in [48]), fueled a growing interest in the concept of streamlined and engineered "minimal" genomes. Soon, researchers began to employ genome-engineering techniques to build such minimal organisms using either bottom-up assembly (i.e., genomes built from scratch, implanted in a cell, and "booted up") [49] or top-down approaches (i.e., the step-wise elimination of portions of a natural genome) [50 51]. Although the effects of genome reduction on the metabolic capacity, stress tolerance, and transcriptomic profile remain largely unknown, these approaches offered an unprecedented opportunity to examine central evolutionary questions outside of the confines of natural genetic variation.

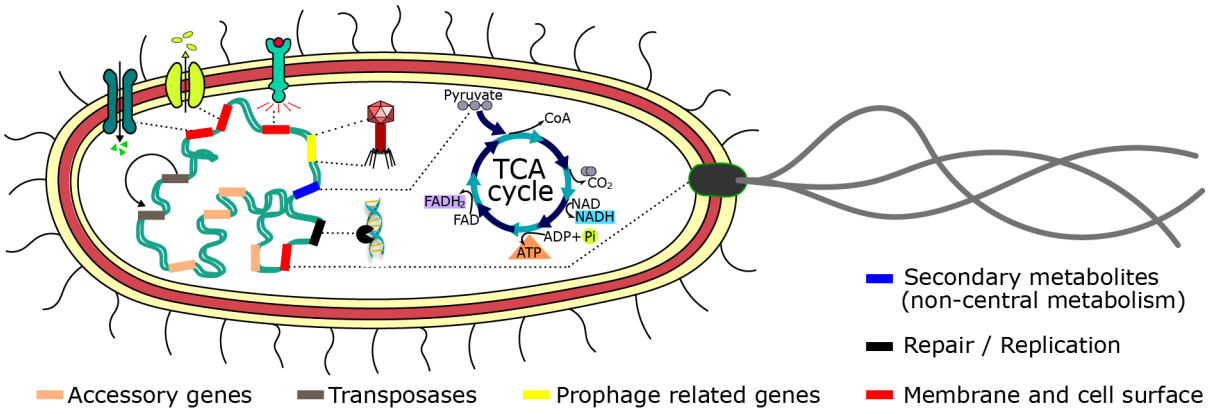

Figure 1.1. Genes commonly targeted for deletion in minimal genome design.

Benefits of clean-genome and minimal-genome strains

In principle, a reduced-genome bacterial chassis could be instrumental to unravel the basic genetics or minimal components necessary to maintain cellular life and provide unprecedented understanding and control of cellular molecular processes [50] [51]. To date, a range of reduced genomes have been engineered from multiple research and industrially relevant strains, most notably, *Escherichia coli*, *Bacillus subtilis* and *Mycobacterium mycoides*, with reductions up to 38.9 % [52], 36.0% [53], and 51.0% [54], respectively (**Fig. 1.2**). "Clean-genomes", as they are sometimes called, remove selfish DNA elements such as transposons (e.g., IS-elements) and prophages [55] [56] [57]. This typically provides greater genome stability and reduced mutation rates [58] [59]. Streamlined genomes further remove nonessential accessory functions, such as genes involved in antibiotic resistance [53], sporulation [53], motility [53], stress resistance [56], and restriction modification systems [60] [57]. Examples like these illustrate how deleting many genes from genomes can provide beneficial strain properties and clues to gene function [55].

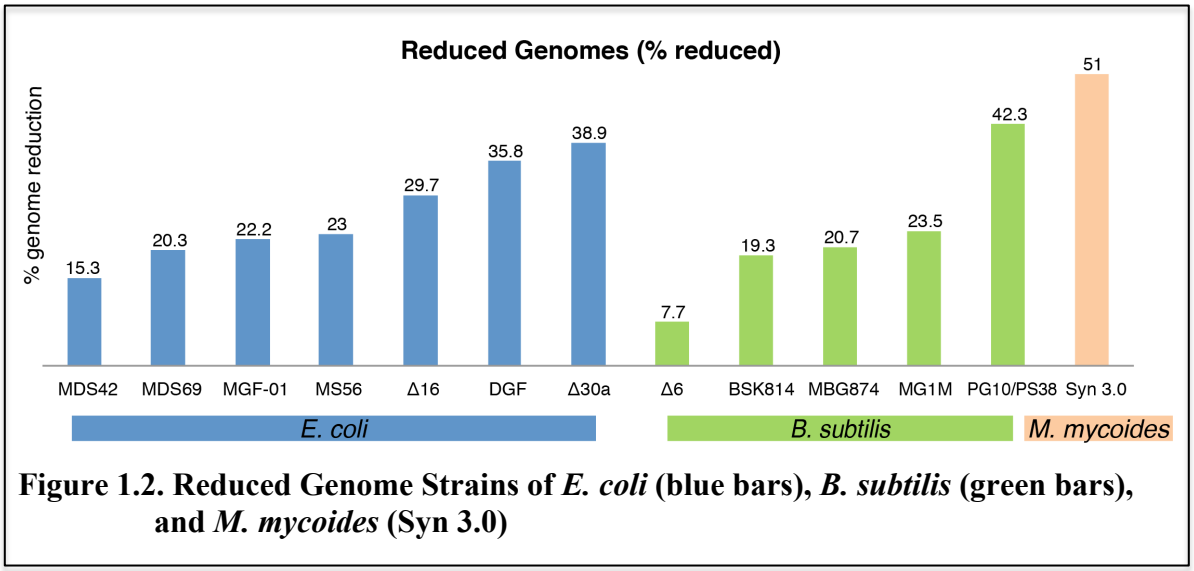

Figure 1.2. Reduced Genome Strains of *E. coli* (blue bars), *B. subtilis* (green bars), and *M. mycoides* (Syn 3.0)

Interestingly, streamlined genomes have exhibited several desirable traits, such as the ability to overexpress 'difficult proteins', improved competence, robust plasmid maintenance, and better tolerance to oxidative stress [61]. It is also possible to pursue genome streamlining to attain other desirable stability features such as reduced autolysis [62 63 64]. Non-detrimental large genome reductions can be superimposed, and in several cases, an improved growth profile has emerged [52,56,65]. Another factor of industrial relevance is the increased biomass or final cell density often observed, even when growth rate was negatively impacted [64 62].

Increased recombinant protein expression is both an objective and attainable outcome of genome reduction efforts [66 56 67 68 69 62 70]. In certain cases, increased productivity is the result of improved plasmid maintenance in reduced genome strains [55 66 58 71 60 64]. In other cases, it is the direct result of the removal of restriction-modification systems or lower mutation rates due to removal of IS-elements and error-prone polymerases [72 73 58].

It is currently debated whether genome reduction positively impacts metabolism in a generalizable way as opposed to idiosyncratic ways that depend on the effects of specific gene deletions. In effect, some reduced-genome strains support improved metabolic efficiency [66 68 62 74 63 64]. Others report no change or even lower performance for reduced-genome cells [62,71]. The changes observed can sometimes be attributed to changes in metabolic capacity, which refers broadly to the availability of precursors (e.g., amino acids), reductive power (e.g., NADPH), energy requirements (e.g., ATP), and gene expression machinery (e.g., ribosomes and RNA polymerases) that are withdrawn from

pools of cellular resources [63 75–77]. Non-essential gene products and secondary metabolites can impose translational or metabolic demands. For example, certain cellular features, such as flagella, impose heavy metabolic burden and thus their removal can free up energy (ATP) and/or reducing power (NADPH) that the deleted strain can now allocate for other functions [63,78]. In fact, evidence suggests that the systematic removal of secondary metabolite producing sequences (e.g., antibiotic biosynthesis operons) and non-essential genes can improve properties over wild-type strains [62].

Obstacles to effective genome streamlining

Genome streamlining remains both time-consuming and challenging. Systematic genome reduction is often accompanied by growth defects, at least in part due to the loss of quasi-essential genes, which are required for robust growth [79–84]. In fact, analysis of the current literature on reduced genomes, reveals that genome reduction most often results in moderate to severe growth impairments. Only in a few cases has increased growth in rich or minimal media been observed: for *E. coli* [85] and *P. putida* [78]. Yet, the general trend is that growth impairments or no significant changes are to be more expected than improved growth phenotypes as a result of minimal genome construction. Most recently, detrimental effects of genome reduction have been suggested to be the result of lack of knowledge of the underlying systems biology [86].

In addition to experimental evidence, comparative genomics and essential gene databases typically guide the selection of deletion targets for genome reduction. One caveat is that essential genes – genes whose deletion causes severe growth impairment or

complete inviability – are not only context-dependent and evolvable, but the fact that a priori knowledge of all essential genes in a cell – its essentialome - does not guarantee the successful engineering of genome-reduced strains [87]. One reason is that deletion of certain gene pairs can be lethal, a phenomena known as "synthetic lethality". Synthetic lethal combinations are difficult to predict during minimal genome construction, particularly when so many gene functions remain unknown. Nonetheless, the combination of modeling and experimental examination can greatly refine the true limits of genome reduction. As an example, an *in silico* tool that integrates genome-scale metabolic and transcriptional models alongside gene essentiality has been shown able to predict the largest "deletable" regions in an *E. coli* genome [88]. Surprisingly, the work suggested a somewhat larger genome dispensability landscape than experimentally tested. In practice, the design trajectory of genome streamlining can dramatically change the outcome or lead to unexpected dead ends. That is, the effect of a adding a new deletion to a cell is contingent on previous deletions. Careful consideration of phenotypic perturbations in reduced-genome cells and how they can be repaired has ameliorated growth defects and resulted in promising streamlined bacterial platforms [79 52,89–91].

The work in this dissertation examines tools and approaches that enable simpler design and engineering of *A. baylyi*. These tools are used to develop and test the effects of genome cleanup and streamlining strategies on this useful microbial "chassis".

Chapter 2: Reduced mutation rate and increased transformability of transposon-free *Acinetobacter baylyi* ADP1-ISx

This chapter is reproduced (with minor modifications) from its initial publication:

Suárez, G.A., Renda, B.A., Dasgupta, A., Barrick, J.E.^ (2017) Reduced mutation rate and increased transformability of transposon-free *Acinetobacter baylyi* ADP1-ISx. Appl. Env. Microbiol. 83: e01025-17. PMID: 28667117

Research Contributions

Suárez, G.A., Renda, B.A. and Dasgupta, A. performed experiments. Suárez, G.A. and Barrick J.E. analyzed data. Suárez, G.A. and Barrick, J.E. wrote the majority of the manuscript.

Acknowledgements

We thank Valérie de Crécy Lagard for the *A. baylyi* ADP1 strain. We also thank three anonymous reviewers for helpful comments. We also thank Daniel Deatherage for sequencing pipeline tips and technical support. We acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for access to high-performance computing resources.

ABSTRACT

The genomes of most bacteria contain mobile DNA elements that can contribute to undesirable genetic instability in engineered cells. In particular, transposable insertion sequence (IS) elements can rapidly inactivate genes that are important for a designed function. We deleted all six copies of IS*1236* from the genome of the naturally transformable bacterium *Acinetobacter baylyi* ADP1. The natural competence of ADP1 made it possible to rapidly repair deleterious point mutations that arose during strain construction. In the resulting ADP1-ISx strain, the rates of mutations inactivating a reporter gene were reduced by 7- to 21-fold. This reduction was higher than expected from the incidence of new IS*1236* insertions found during a 300-day mutation accumulation experiment with wild-type ADP1 that was used to estimate spontaneous mutation rates in this strain. The extra improvement appears to be due in part to eliminating large deletions caused by IS*1236* activity, as the point mutation rate was unchanged in ADP1-ISx. Deletion of an error prone polymerase (*dinP*) and a DNA damage response regulator (*umuDAb*) from the ADP1-ISx genome did not further reduce mutation rates. Surprisingly, ADP1-ISx exhibited increased transformability. This improvement may be due to less autolysis and aggregation of the engineered cells compared to wild type. Thus, deleting IS elements from the ADP1 genome led to a greater than expected increase in evolutionary reliability and unexpectedly enhanced other key strain properties, as has been observed for other clean-genome bacterial strains. ADP1-ISx is an improved chassis for metabolic engineering and other applications.

IMPORTANCE

Acinetobacter baylyi ADP1 has been proposed as a next-generation bacterial host for synthetic biology and genome engineering due to its ability to efficiently uptake DNA from its environment during normal growth. We deleted transposable elements from the ADP1 genome that are capable of copying themselves, inserting into other genes, and thereby inactivating them. The resulting 'clean-genome' ADP1-ISx strain exhibited larger reductions in the rates of inactivating mutations than expected from spontaneous mutation rates measured via whole-genome sequencing of lineages evolved under relaxed selection. Surprisingly, we also found that IS element activity reduces transformability and is a major cause of cell aggregation and death in wild-type ADP1 grown under normal laboratory conditions. More generally, our results demonstrate that domesticating a bacterial genome by removing mobile DNA elements that have accumulated during evolution in the wild can have unanticipated benefits.

INTRODUCTION

Acinetobacter baylyi ADP1 is a nonpathogenic Gram-negative soil bacterium that has been used as a platform for metabolic engineering and synthetic biology due to its high natural transformability [92–95]. ADP1 naturally catabolizes aromatic compounds and can detoxify inhibitory molecules that are liberated during degradation of lignocellulosic feedstocks [29,96]; and it can be engineered for the production of biofuel components and high-value compounds, including wax esters and triacylglycerides [26,28]. ADP1 has also been used to construct cell-based biosensors for salicylate and components of crude oil [31,97]. Facile engineering of the ADP1 genome to edit and add to metabolic and sensing pathways is made possible by its constitutive expression of a competence apparatus that enables it to efficiently uptake DNA from its environment in a non-sequence-specific manner [21]. ADP1 can also transform fragmented and damaged DNA or mutagenized PCR products to introduce sequence variation at specific sites in its genome for optimizing engineered functions [8,98].

Insertion sequence (IS) elements are minimal transposons that are widespread in bacteria [99]. They contribute to mutagenesis by encoding transposases that move or copy their DNA sequences to new locations in a genome and by acting as long repeated sequences that can mediate homologous recombination events that lead to deletions and other large-scale genome rearrangements [100,101]. IS elements can be significant sources of genetic instability in engineered bacterial cells. For example, they are often the dominant

source of spontaneous mutations that rapidly inactivate the heterologous expression of costly recombinant proteins from plasmids [44,58].

The genome of *A. baylyi* ADP1 contains a total of six IS elements, all of the same IS*1236* type [102,103]. IS*1236* elements are members of the IS*3* family [104]. They operate by a copy-paste mechanism and insert randomly without any strong target site sequence bias. Five of the six IS*1236* copies have identical sequences. Two of these are located in close proximity on the chromosome, forming a composite transposon (Tn*5613*) that can also mobilize two hypothetical genes located between the IS copies [105]. The sixth copy (IS*1236ψ*) has evolved to be non-autonomous. It shares only 82% nucleotide identity with the other copies and has a frameshift mutation in its transposase gene [106].

We and others have observed that a considerable proportion of the mutations that occur in *A. baylyi* ADP1 are due to IS*1236* activity. This transposable element family was initially characterized due to the prevalence of IS*1236* insertions among mutations inactivating certain protocatechuate catabolism genes [103]. IS*1236* elements constitute 25-40% of all new mutations that accumulate during long-term adaptive laboratory evolution experiments with ADP1 [107,108], including mutations in competence genes that often arise during laboratory culture and reduce or eliminate the transformability of this strain [108,109]. The DNA cleavage activity associated with IS*1236* elements may also contribute to certain gene amplification events [106,110].

In this study, we sequenced the genomes of *Acinetobacter baylyi* ADP1 that had been evolved under relaxed selection for 300 days and found that IS*1236* elements were directly responsible for 26% of all spontaneous mutations. Therefore, to improve the

genetic stability of ADP1, we engineered an IS-less version of this strain (ADP1-ISx) by deleting all six IS*1236* copies. Characterization of the final strain revealed that ADP1-ISx exhibited a much larger reduction in mutation rates than expected, and also significantly greater cell yields and transformation frequencies. These improved traits in the 'clean-genome' ADP1-ISx strain make it an improved platform for bioengineering applications that take advantage of its metabolic versatility and high naturally transformability.

MATERIALS AND METHODS

Strains and culture conditions

Acinetobacter baylyi ADP1 cultures were grown in LB or minimal succinate (MS) media at 30°C with orbital shaking, as previously described [19]. Media were supplemented with μg/ml kanamycin (Kan), 60 μg/ml spectinomycin (Spec), or 200 μg/ml 3′-azido-2′,3′-dideoxythymidine (AZT) as specified. Engineered ADP1 strains were archived at –80°C in 15-20% (v/v) glycerol. Prior to all experiments, cell stocks were thawed on ice, diluted 1:1000 into the appropriate media, and grown overnight. Then, these revived cultures were diluted 1:1000 and grown for an additional 24±1 h for preconditioning to the culture conditions of an experiment. Cell dilutions for plating on agar were made in sterile saline (0.85% w/v NaCl).

Mutation accumulation (MA) experiment

The MA lines were generated in prior work [19]. Briefly, 18 ADP1 lineages were passaged through 300 daily single-cell bottlenecks by picking an arbitrary colony from LB agar, streaking out to single colonies, and allowing regrowth for 24 h before selecting

the next colony to transfer. This procedure gives ~25 cell divisions (generations of regrowth) each day for an estimated total of ~7,500 generations of evolution over the course of the entire MA experiment.

Genome sequencing

DNA was isolated from evolved clones from the MA experiment and genome-edited clones from the construction of ADP1-ISx and then prepared for sequencing as described previously [108]. These samples were sequenced to >100× coverage using the Illumina HiSeq 2500 platform at the University of Texas at Austin Genome Analysis and Sequencing Facility (GSAF). Read files were analyzed by using the *breseq* computational pipeline (v0.28.0) [111,112] to predict mutations, including IS element insertions and other types of structural variation, relative to the ADP1 reference genome (GenBank: NC_005966.1) [23]. FASTQ files for each sequenced strain have been deposited in the NCBI Sequence Read Archive (accession SRP074541). When counting mutations in the MA experiment we removed genome differences that we previously identified as being in our original lab strain [19], and we also did not include an insertion of a single G that was present in 15 of the 17 sequenced genomes at a site 74 base pairs upstream of the *atpI* gene because it was most likely already present in the original culture that was used to found these lineages. Statistical tests were performed in R (v3.3.2) [113]. The final mutation predictions are provided in **Table S1**.

Genome modification

Enzymes were purchased from New England Biolabs (Ipswich, MA). Phusion polymerase was used for all PCRs. Primer sequences are provided in **Table 2.1**. The *tdk*/Kan^R dual-selection cassette was PCR amplified from the *per* gene knockout strain from the ADP1 single-gene deletion collection [114] using primers P1-F and P2-R. For each IS or gene deletion, ~1000 base pairs of flanking sequence on each side of the deleted region were amplified in two additional PCR reactions. Primers P3-F and P1′P4-R were used for the upstream region, and primers P2′P5-F and P6-R were used for the downstream region. PCR products were cleaned up using the GeneJET Purification Kit (Thermo Scientific). The two flanking region products and the *tdk*/Kan^R cassette were joined by overlap PCR to create the 'knockout cassette'. Two PCR products for the flanking regions were joined in a separate overlap PCR reaction to produce the 'rescue cassette' used for scarless *tdk*/Kan^R cassette removal. To add overlap in this case, the upstream region was amplified with primers P3-F and P5′P4-R, and the downstream region was amplified with primers P4′P5-F and P6-R. Each type of overlap PCR amplification began with 15 temperature cycles with only the templates present (no primer addition), followed by 20 further cycles with the external primers P3-F and P6-R. The final amplified cassettes were digested with 40 U DpnI by adding 2 µL of enzyme into each 50 µL PCR reaction and incubating for 1 h at 37°C. Then, they were purified again using the GeneJET PCR Purification Kit before transformation.

Each genome modification involved two transformation steps. First, 70 µL of ADP1 cells from a preconditioned LB culture of the strain being edited were added to 1

ml of LB containing 100 ng of knockout cassette DNA. After overnight growth (16-24 h), 100 μL of the transformation culture and 100 μL of a 1:10 dilution were plated on LB-Kan plates and 100 μL of a 10^6 dilution were plated on LB to monitor transformation frequencies. In most cases, plating 100 μL from a 1:10 dilution of a transformation culture resulted in ~30 colonies on LB-Kan. In the second step, the integrated *tdk*/Kan^R cassette was replaced using the corresponding rescue cassette PCR product with the same transformation procedure, except plating on LB-AZT plates to select for removal of the *tdk*/Kan^R cassette. Each step was typically carried out with three or more replicates so that independently derived clones could be screened for success. Whole-cell PCR was used to confirm successful integration and loss of the *tdk*/Kan^R cassette at each step.

Each IS element was deleted one at a time using this procedure, except for the two IS*1236* elements in Tn*5613* (#2 and #3), which were deleted simultaneously. This double IS deletion also removed two putative genes of unknown function (*ACIAD0955* and *ACIAD0956*) located between the IS elements and was designed to preserve a stop codon in an adjacent gene of unknown function (*ACIAD0959*) that overlaps IS*1236* #3.

The *dinP* and *umuCDAb* deletions were constructed in the repaired ADP1-ISx strain background. The double mutant was made by adding the *umuDAb* deletion to the Δ*dinP* strain. The Δ*umuDAb* single mutant produced small colonies on LB agar and did not reach saturation in LB liquid after 24 h of growth. Its growth defect appears to be due to a secondary mutation that occurred during our strain construction as Δ*umuDAb* strains constructed by others have been reported to exhibit normal growth [115]. The Δ*dinP* and Δ*dinP* Δ*umuDAb* strains that we constructed grew normally.

qPCR monitoring of IS deletion

We used quantitative PCR (qPCR) to monitor the relative copy number of IS*1236* elements during ADP1-ISx strain construction. Genomic DNA (gDNA) was purified from at least three of the clones confirmed by PCR to have a successful deletion of each targeted IS copy. The Qubit dsDNA BR Assay Kit (Invitrogen) was used to determine DNA concentrations in these samples.

qPCR reactions were set up with gDNA (1.1 ng/μL), IS-F and IS-R primers (0.5 μM each), and SYBR Green Dye Real-Time PCR Master Mix (Applied Biosystems). These primers amplify a 119-bp product common to all six IS*1236* elements (**Table 2.1**). More cycles of PCR amplification (R_n) were required to surpass an arbitrary signal threshold in each successive strain that removed additional IS*1236* elements from the ADP1 genome. Clones with a higher R_n value were used to continue the deletion procedure.

Growth curves

Growth curves for wild-type ADP1 and ADP1-ISx were initiated by making 1:1000 dilutions of cultures preconditioned in LB or MS into 50 mL of the same medium in 250 mL Erlenmeyer flasks. The optical density of samples removed from these cultures was measured at 600 nm (OD600) to monitor growth. These assays were carried out in triplicate. Time points at which the mean of the OD600 values across replicates was less than 0.25 were used for nonlinear least-squared fitting to a model with exponential growth rate and lag time parameters using R (v3.3.2) [113]. Differences in growth rates or lag times between two strains were evaluated by simultaneously fitting

the OD600 data for both strains to a model that had allowed just one global value for this parameter for both strains and then examining the significance of an added offset that allowed for per-strain variation in that parameter.

Reversion of *cyoB* and *rpoD* mutations

Regions of the wild-type ADP1 genome consisting of ~1,000 base pairs upstream and ~1,000 base pairs downstream of the *cyoB* or *rpoD* mutations were amplified by PCR. These PCR products were DpnI digested and gel purified. Then, 250 ng of each PCR product were transformed into the same culture of ADP1-ISx using standard conditions. After 12 h of growth, 100 µL of this culture was transferred into 10 ml fresh LB. After 6 h of growth, 100 µL was transferred again into 10 ml LB. After a final 6 h of growth, 100 µL of a 10^6 dilution was plated on LB agar. PCR and Sanger sequencing showed that all twelve of the large colonies picked from this LB plate had reverted to wild-type sequences for both genes. After whole-genome sequencing to confirm that it had not accumulated any new secondary mutations, one of these clones was designated ADP1-ISx and used in all further experiments.

Gene inactivation mutation rate assays

Three pairs of strains derived from ADP1-ISx and wild-type ADP1 were constructed, each with the Kan^R/*tdk* cassette inserted at a different genomic site using the methods described above. For each of these strains, thirty independent 40 µl LB cultures, each containing ~500 colony-forming units (CFUs) from a dilution of a preconditioned culture, were started in 96-well microplates. These plates were sealed with adhesive foil

and incubated at 30°C with orbital shaking at 250 r.p.m. Following overnight growth, the entire volume of twenty-four of the cultures for each of the six strains were independently plated on LB-AZT plates. A 10^6 dilution of the remaining six 40 µL cultures of each strain was plated on nonselective LB plates. After incubation at 30°C for 30 h, colonies on selective and non-selective plates were counted. Mutation rates were calculated from these counts using FALCOR [116].

To compare the types of mutations leading to AZT-resistance in each strain, one colony was picked from each selective plate in the fluctuation tests and grown overnight in liquid LB. We then used PCR with primers P1-F and *tdk*-R to amplify the *tdk* portion of the Kan^R/*tdk* cassette from the genome of each of these mutants.

Point mutation rate assays

We utilized a spectinomycin resistance gene in which the eighth codon was mutated from GAA to the TAA stop codon to measure point mutation rates. To create this $Spec^R$-E8* mutational reporter, we amplified the spectinomycin resistance gene from pIM1463 [117] in two halves with primers that introduced the G to T mutation in a region common to each PCR product, and then we stitched the two pieces together in a round of overlap PCR. We inserted this marker into the *A. baylyi* genome in 'region 1', a site where we had inserted a Kan^R cassette in a previous study [19]. Specifically, strains ADP1-$Spec^R$-E8* and ADP1-ISx-$Spec^R$-E8* were created by first integrating a Kan^R/*tdk* cassette in region 1 and then using the mutant $Spec^R$-E8* DNA assembled with regions of flanking homology as a rescue cassette.

The point mutation rate was assessed for ADP1-Spec^R-E8* and ADP1-ISx-Spec^R-E8* using Luria-Delbrück fluctuation tests in LB. Preconditioned cultures were diluted and used to inoculate 1 mL test tube cultures with ~30-50 cells. After growth for ~18-24 h, cells from 24 of the tubes for each strain were pelleted via centrifugation and the entire amount was plated on LB-Spec agar. These selective plates were incubated at 30°C for 4 days before counting colonies. The other six replicates were diluted 10^6 in sterile saline plated on nonselective LB agar and incubated overnight before counting. Point mutation rates were calculated and compared using the rSalvador R package (version 1.7) [118,119].

Transformation frequency

Six replicate 0.5 mL transformation reactions were prepared with 250 ng/mL of DNA consisting of the Kan^R/*tdk* cassette PCR amplified with ~1000-bp flanking homologies targeting site 2 or the unmutated Spec^R cassette amplified with ~1000-bp flanking homologies targeting site 4. At the same time, dilutions from every 0.5 mL transformation mix were plated onto LB agar to acquire CFU counts corresponding to each. Negative controls with DNA but no cells were included to rule out possible contamination issues. All samples and transformation mixes were incubated for 16 h. Dilutions of each transformation were plated on LB-Kan and LB agar, and transformation frequencies were calculated by taking the ratios of CFUs after overnight incubation at 30°C on selective versus nonselective plates.

Extracellular DNA (eDNA) concentration

Eight 5 mL cultures in MS medium were inoculated with a 1:1000 dilution of preconditioned cells and grown for 48 h. eDNA measurements were made using the Qubit high-sensitivity (HS) kit on the supernatant from 1 mL samples from these cultures after centrifuging for 10 min at 21,130 r.c.f.

Live/dead staining

Cells from preconditioned cultures grown for 24 h in MS medium were analyzed using the Live/Dead® BacLight™ Bacterial Viability and Counting Kit (Molecular Probes). For each biological replicate, 10,000 events were captured on a BD LSRFortessa Flow Cytometer and analyzed with FlowJo (v10.1). We analyzed five biological replicates in LB and three biological replicates in MS media for each strain. Live/dead gating was determined using dead cell controls prepared by incubating cultures with 70% (v/v) isopropanol for 1 h at room temperature.

RESULTS AND DISCUSSION

Spontaneous mutation rates in *A. baylyi* ADP1

We measured genome-wide spontaneous mutation rates in ADP1 by sequencing 17 strains from a previous mutation accumulation (MA) experiment [19]. Each of these clonal isolates represents the endpoint of a lineage that evolved for ~7,500 generations (cell doublings) over 300 days of passaging colonies on LB agar. The filtering effect of natural selection is greatly reduced in MA experiments because every lineage experiences a single-cell bottleneck when a random colony is selected each day for propagation. This

design enables genetic drift to dominate over selection and causes the rates at which mutations appear in MA lineages to be very close to the rates at which they spontaneously arise in a genome [120].

We found a total of 66 mutations across all 17 clones, with zero to nine mutations in each individual clone (**Fig. 2.1A**). Despite this wide range, the distribution of mutation counts across these independent lineages was still compatible with no difference in the underlying genomic mutation rate that each experienced ($P = 0.45$, Kruskal-Wallis test). The overall estimated mutation rate for ADP1 is 0.52 mutations per genome per 1000 generations (0.40–0.66, Poisson 95% confidence interval).

A majority of the mutations observed in the MA lines were base substitutions ($n = 36$). The estimated total base substitution mutation rate is 0.79×10^{-10} per base pair per generation ($0.55–1.09\times10^{-10}$, Poisson 95% confidence interval). This value is similar to that found in other γ-proteobacteria, including *Escherichia coli* [121], *Vibrio* species [122], and *Pseudomonas aeruginosa* [123], for which estimates from MA experiments range from 0.7 to 2.1×10^{-10} substitutions per base pair per generation. Other mutational signatures in ADP1 were also similar to those observed in other bacterial species. Transitions were more common than transversions, and G:C→A:T base pair substitutions were the dominant type of point mutation (**Fig. 2.1B**).

Of the 29 base substitutions in protein coding regions, 8 were synonymous. If these mutations were redistributed purely by chance in coding portions of the ADP1 genome while preserving the observed number of each type of base pair substitution (e.g., 20 G:C→ A:T mutations), then 13.8% to 37.9% of these mutations would be expected to

be synonymous (95% confidence interval). Thus, the observation of 27.6% synonymous mutations is compatible with greatly reduced selection for fitness during the experiment.

Of the 66 total mutations, 7 were insertions of IS*1236* elements at new locations in the genome and 10 were deletions with at least one end adjacent to an IS*1236* element (**Fig. 2.1A**). These 17 IS-mediated mutations were spread throughout 12 of the 17 sequenced clones, with a single event per genome except that one clone had five of these mutations and another had two. This proportion of IS-mediated mutations (17/66) was not significantly different from that observed for a set of largely beneficial mutations (30/75) found in a previous adaptive evolution experiment [19] (two-tailed Fisher's exact test, $p = 0.073$), implying that, when considered together, IS element mediated mutations are not systematically much more or less beneficial than other types of mutations in *A. baylyi*. Four of the IS*1236* mediated deletions were recombination events between the two IS elements flanking Tn*5613* that eliminated its two cargo genes of unknown function. These deletions were also commonly observed in the adaptive evolution experiment [108]. The remaining 13 non-point mutations included 6 indels of 50 or fewer base pairs, 5 larger deletions, and two larger duplications (**Fig. 2.1A**).

ADP1-ISx strain construction

To eliminate IS elements as a source of mutations that could inactivate engineered DNA constructs, we deleted all six of the IS*1236* elements in the ADP1 genome in five sequential cycles of genome editing to create strain ADP1-ISx (**Fig. 2.2A**). To avoid the unwanted spread of new IS*1236* copies into unknown regions of the genome during strain

construction, we used qPCR to verify that IS*1236* copy number in the genome decreased after each deletion (**Fig. 2.2B**).

Whole-genome sequencing of candidate ADP1-ISx strains revealed that they were free of IS*1236* elements but had all accumulated two secondary mutations (**Table S1**). The altered genes were *rpoD*, which sustained an in-frame deletion of three base pairs in a $(AAG)_3$ repeat, and *cyoB*, which acquired a nonsynonymous point mutation. RpoD is the ADP1 homolog of σ^{70}, the major housekeeping sigma factor [124]. CyoB is a subunit of the cytochrome *bo* terminal oxidase complex that is involved in respiration [125].

The ADP1 genome editing procedure involves repeated transformations, growth of liquid cultures in LB, and growth of colonies on LB agar in the presence of kanamycin or azidothymidine [92]. Secondary mutations can accumulate by chance if they happen alongside the desired edits whenever a single colony is picked during this procedure (much like in the MA experiment). Alternatively, given the important roles of *rpoD* and *cyoB* in aerobic growth, it is possible that one or both of these mutations might have been favored during some or all of these culture conditions. In particular, the *cyoB* mutation may have been beneficial in response to repeated kanamycin exposure [126].

Under normal laboratory conditions an ADP1-ISx candidate strain with these two mutations exhibited greatly compromised growth (**Fig. 2.3**). It had significantly longer lag times and doubling times in liquid cultures compared to wild-type ADP1 in both LB broth and MS defined medium ($P < 10^{-6}$, for two-tailed tests on each parameter in both media as described in the **Methods**). For example, the doubling time of the ADP1-ISx-*rpoD-cyoB* strain was 65 minutes in LB compared to 35 minutes for wild-type ADP1.

Therefore, we reasoned that it might be possible to repair one or both of these mutations by transforming PCR products containing the wild-type versions of these two genes to an ADP1 culture and then passaging these cells. After <30 generations of serial transfer and regrowth in LB we found that every large colony that we sequenced had reverted to wild-type versions of both genes. Whole-genome sequencing verified that one of these colonies now had no mutations in its genome aside from the five planned deletions of IS*1236* elements. Growth curves of this final ADP1-ISx strain (**Fig. 2.3**) showed that its doubling time during exponential growth was now indistinguishable from wild-type ADP1 in both LB and MS ($P = 0.94$ and $P = 0.82$, respectively). Interestingly, ADP1-ISx was able to grow to saturation more rapidly because it had a significantly shorter lag time than wild-type ADP1 in both LB and MS ($P = 0.011$ and $P < 10^{-6}$, respectively).

Reduced rate of inactivating mutations in ADP1-ISx

We next used a forward mutation assay to determine if mutation rates were reduced in the engineered ADP1-ISx strain. Specifically, we measured the rates of loss-of-function mutations in a *tdk* reporter gene when it was integrated at several different chromosomal locations in each strain background (**Fig. 2.4A**). Loss of function of this gene gives resistance to killing by azidothymidine (AZT^R). We found ~21-fold (Site 1), ~7-fold (Site 2) and ~13-fold (Site 3) reduced mutation rates in ADP1-ISx compared to wild-type ADP1 (**Fig. 2.4A)**.

Variation among the mutation rates measured at each of these three sites is consistent with greater stabilization of the ADP1-ISx chromosome at locations that are

nearer to active IS*1236* elements in wild-type ADP1. The largest mutation rate reduction was found at Site 1, which is only 1.3 kb away from Tn*5613*. Site 3 is located 76.9 kb from the nearest IS element and has an intermediate level of reduction in mutation rates. Site 2, which exhibited the smallest mutation rate reduction, is also located very near an IS element, but it is the mutated IS*1236φ* copy. Therefore, this result provides further evidence that this element is inactive. Taking this into account, Site 2 is actually the furthest of the three sites from the nearest active IS element, which is 910.4 kb away.

Overall, the rates of inactivating mutations in wild-type ADP1 at different sites in the chromosome varied by around 3-fold, and these mutation rates were highest at the two sites that were within 100 kb of active IS-elements. This range of variation is broadly consistent with previous studies that have looked at the proportion of inactivating mutations caused by insertions of IS*1236* and Tn*5613* at various genetic loci in ADP1 [103,105]. However, all of the measurements comparing ADP1-ISx to wild-type ADP1 found much larger decreases in mutation rates than the ~1.3-fold expected from the overall contribution of IS elements to mutagenesis found in the MA experiment where they accounted for just 26% of all spontaneous mutations. Even considering that IS insertions are more likely to disrupt a gene than point mutations, inactivating mutations were much rarer than we expected in ADP1-ISx.

Unchanged point mutation rate in ADP1-ISx

In order to understand if lower rates of point mutations in ADP1-ISx contributed to the greater than expected improvement in its genetic stability compared to wild-type

ADP1, we next performed a reverse mutation assay (**Fig. 2.4B**). Specifically we inserted a spectinomycin resistance gene with a stop codon introduced near the beginning of its reading frame into the chromosome of each strain at Site 4 (**Fig. 2.2A**). Selecting for spectinomycin resistant mutants is expected to be specific for measuring the rates of base substitutions that restore gene expression by changing the stop codon to a sense codon. The mutation rate estimated for ADP1-ISx (3.8×10^{-9} per cell per generation) for this assay was slightly higher than that found for wild-type ADP1 (2.7×10^{-9} per cell per generation), but this difference in mutation rates was not statistically significant (P = 0.054, likelihood ratio test) [118]. Thus, point mutation rates appear to be unchanged or, if anything, slightly higher in ADP1-ISx.

Altered mutational spectrum in ADP1-ISx

Next, we compared the types of inactivating mutations that occurred in ADP1-ISx versus wild-type ADP1 (**Fig. 2.4C**). For each of the six test strains, we analyzed 24 independent AZT^R mutants isolated during the fluctuation assays by attempting to amplify a 767-bp PCR product containing the *tdk* gene from their genomes. Insertion of a new IS*1236* copy within the *tdk* gene would increase the size of this PCR product to ~2,000 bp. Point mutations and small insertions or deletions in the *tdk* gene would lead to no visible change in PCR product size. Larger deletions that removed a portion of the *tdk* gene and adjacent genome regions extending past the primer binding sites would lead to no amplification product. To corroborate these PCR results, we further tested for whether function of the Kan^R gene located adjacent to *tdk* in the reporter cassette was maintained

in each mutant. If a strain with no PCR product was also kanamycin sensitive, then it was further evidence that a larger deletion that overlapped both the *tdk* gene and the adjacent Kan^R marker gene had occurred. If, on the other hand, a strain with no PCR product remained kanamycin resistance, then we classified the mutation as an "ambiguous" diagnosis of the mutation. This situation could indicate that there was a deletion overlapping the *tdk* gene but not the Kan^R marker gene, or it could be the result of a PCR reaction that failed for some other reason.

As expected, we found no evidence of IS insertions in the ADP-ISx strains, though they were responsible for 12.5% to 29.2% of the mutations in wild-type ADP1 depending on where the reporter gene was integrated into the chromosome. For Site 1 and Site 2, a large fraction of the inactivating mutations in wild-type ADP1 resulted in no PCR band: 58.3% for Site 1 and 70.8% for Site 2. A majority of these mutants were also kanamycin sensitive (10/14 for Site 1 and 12/17 for Site 2), which is consistent with these mutants having large deletions that overlap both the *tdk* and Kan^R genes. The spectrum of mutations shifted markedly in ADP1-ISx. Only point mutations were found at Site 2 in ADP1-ISx and 87.5% of mutations at Site 1 were point mutations. Thus, deletion of IS elements from ADP1 resulted in a lower rate of IS insertions, as expected, but also fewer large deletions, which are apparently mediated by the action of IS elements. This result could potentially explain the greater-than-expected reduction in mutation rates in ADP1-ISx.

Site 3 exhibited a different change in mutational signature. Here, there were fewer mutants with no PCR band in wild-type ADP1 (41.7%), and the proportion of these

mutants was similar in ADP1-ISx (54.2%). Again, these mutants were mostly kanamycin sensitive, 7/10 for wild-type ADP1 and 10/13 for ADP1-ISx. These results are consistent with large deletions occurring at this site via a mechanism that is not directly dependent on IS activity. We found a 14/15 bp match for a sequence within the *aph* (Kan^R) gene to a region 3,108 bp upstream of the *tdk*/Kan^R cassette insertion site that is a candidate for homology that might mediate a deletion hotspot here [127]. However, it remains unclear why there is an overall decrease in the mutation rate at Site 3 without a commensurate shift in the mutational spectrum away from IS-related mutations.

Increased transformability with reduced autolysis in ADP1-ISx

As we constructed strains for testing mutation rates by transforming reporter cassettes, we noticed that this process appeared to be more efficient in ADP1-ISx. Directly testing for a change showed that ADP1-ISx exhibited significantly increased transformation frequencies for PCR products containing the Kan^R/*tdk* cassette (into Site 2) or a $Spec^R$ gene (into Site 4) (**Fig. 2.5A**). ADP1-ISx was 7.6-fold and 3.3-fold more transformable than wild-type ADP1 in the Site 2 and Site 4 assays, respectively.

We also observed a slight but significant increase in the final optical density (OD) of cultures of ADP1-ISx (**Fig. 2.5B**) in both LB and MS media ($P = 0.00098$ and $P = 0.0028$, two-tailed Welch's *t*-tests comparing ODs after 18 h in LB and MS, respectively). In further exploring this difference, we found that there were ~4 times as many colony-forming units (CFUs) in saturated LB cultures of ADP1-ISx as there were for wild-type ADP1 (**Fig. 2.5B**). Additionally, when these cultures were left unshaken on

a bench, wild-type ADP1 settled to a pellet in the bottom while ADP1-ISx cells remained in suspension for days (**Fig. 2.5C**). We have seen differences in cellular aggregation in evolved ADP1 strains previously, though it was increased aggregation in this case [19]. Taken together, these observations seemed to indicate that there were either more viable cells, less aggregation of those cells (leading to more CFUs per cell), or both in ADP1-ISx cultures.

One possible explanation for the reduced viable cell counts and transformability of wild-type ADP1 compared to ADP1-ISx could be that deleting IS*1236* elements decreases autolysis. Genomic DNA released from dead cells would be expected to compete with the exogenous DNA added in transformation assays for uptake into cells. In line with this hypothesis, we found that the concentrations of extracellular DNA that accumulated in cultures grown in MS medium were higher for wild-type ADP1 (154 ng/mL) than they were in ADP1-ISx (63.2 ng/mL) at 24 h, and this difference became even more pronounced after 48 h (**Fig. 2.5D**). Live/dead staining found high amounts of autolysis in normal cultures of wild-type ADP1 grown in LB broth while ADP1-ISx cultures exhibited much less cell death (**Fig. 2.5E**). The same trend was found in MS medium, though the fraction of cells staining as dead was lower for both strains in these conditions: 6.4 ± 4.6% for wild-type ADP1 and 1.8 ± 1.4% for ADP-ISx (95% CI). Thus, reduced autolysis of ADP1-ISx may contribute to its improved transformability.

It is also possible that the improved transformability of ADP1-ISx is due to other factors. For example, the activity of IS*1236* elements could inhibit successful transformation after exogenous DNA is inside a cell. This effect could be direct, if the

IS*1236* transposase interacts with or inactivates DNA structures that are intermediates involved in successful integration into the chromosome. Alternatively, IS*1236* activity could indirectly modulate DNA repair pathways that are also required for transformation [128], by drawing the necessary factors away to sites of transposition, for example. Increased aggregation of cells in a wild-type ADP1 culture compared to an ADP1-ISx culture related to the settling phenotype might also reduce transformability by restricting the surface area of ADP1 cells that is accessible to added DNA for uptake.

Importantly, increased aggregation might complicate our measurements of transformation frequencies and mutation rates if one CFU contains multiple (and more) cells for wild-type ADP1 compared to ADP1-ISx. Only one cell in an aggregate needs to be transformed or mutated for it to form a colony on a selective plate, yet it is counted as a single CFU on a nonselective plate regardless of the number of cells in the aggregate. The consequence is that what are normally assumed to be "per-cell" transformation and mutation rates will be overestimated for a strain exhibiting more aggregation because they actually reflect "per-aggregate" rates. Essentially, there are multiple chances (one for each cell) for the aggregate as a whole to give a signal of transformation or mutation.

Considering the potential effect of aggregation on our measurements, it is clear that it could not explain our finding that ADP1-ISx is more transformable than wild-type ADP1. The opposite relationship, with ADP1 appearing to be more transformable, would be expected if aggregation were the only factor. By contrast, an aggregation artifact could potentially explain why there was a larger decrease in the apparent per-cell inactivating mutation rates estimated for ADP1-ISx relative to wild-type ADP1 than was expected

from the overall contribution of IS*1236* to mutagenesis (**Fig. 2.4A**), including the large decrease at Site 3 without much change in the mutational spectrum (**Fig. 2.4C**). However, we would also expect to find a commensurate difference in our measurements of point mutation rates between wild-type ADP1 and ADP1-IS, and we do not (**Fig. 2.4B**). Therefore, we conclude that any change in aggregation between wild-type ADP1 and ADP1-ISx does not explain the dominant trends we see in these measurements.

Deleting error-prone polymerases does not further reduce the mutation rate

Deletion of the three stress-induced error-prone polymerases Pol II, Pol IV, and Pol V, from IS-less *E. coli* strain MDS42 has been shown to reduce its inactivating mutation rate by approximately 50% [129]. We investigated whether a similar improvement was possible in ADP1-ISx. ADP1 encodes homologs of two of these *E. coli* polymerases: Pol IV (*dinP*) and Pol V (*umuDC*). However, the polymerase function of the *umuDC* complex is no longer intact. ADP1 *umuC* is a pseudogene interrupted by a transposase fragment derived from an ISEhe3-like element [115]. The size and regulation of the ADP1 *umuD* gene differs from *E. coli,* and this distinct protein subfamily (designated *umuDAb*) acts as a regulator of the DNA damage response in *Acinetobacter* species [115,130,131].

We created ADP1-ISx strains with *dinP* and *umuDAb* deleted, singly and in combination. No significant difference in the rate of inactivating mutations in the *tdk* gene integrated into the chromosome at Site 2 was found among these three deletion strains (**Fig. 6**). To test whether there was a difference in mutagenesis under conditions in which error prone polymerases expression might be induced to a higher level, we repeated this assay in the presence of subinhibitory concentrations of nalidixic acid.

Nalidixic acid causes double-strand breaks and thereby induces the expression of error-prone polymerases via the SOS response in *E. coli* and other bacteria [132]. We found that nalidixic acid increased the rate of mutations in the Δ*umuDAb* mutant and ADP1-ISx to the same extent. Thus, deleting *dinP* and *umuDAb* appears to have no further benefit with respect to further reducing the mutation rate of ADP1-ISx.

Comparison to other clean-genome bacterial strains

Several other studies have deleted selfish DNA elements and nonessential genes from bacterial genomes to enhance their biotechnology performance characteristics [62–64,67,69,85,133–135]. One comprehensive effort has been the development of the streamlined *Escherichia coli* strain MDS42 [133]. MDS42 was derived from *E. coli* K12 by deleting all active IS elements, all prophage remnants, and many nonessential genes from its chromosome. The rates of mutations inactivating a chromosomal reporter gene are lower in MDS42 than its progenitor strain [133], and these mutation rates were further reduced in subsequent work by deleting all three error-prone polymerases from its genome [129]. Genetic stability is improved in MDS42 to the point that plasmids encoding highly toxic proteins that are rapidly inactivated in wild-type *E. coli* can be propagated in this strain [58]. MDS42 also exhibited an unanticipated increase in electroporation efficiency. The exact reason for this improvement is unknown, as many genes were deleted in this strain, but it is thought to result from the deletion of fimbriae that might obscure access of DNA to the cell surface.

Pseudomonas putida EM383, a streamlined chassis strain derived from KT2440, also showed enhanced properties after the deletion of flagellar genes, four prophages, and both of its active transposons (Tn*7* and Tn*4652*) [64]. EM383 exhibits increased retention of functional plasmids during prolonged culture and has improved growth characteristics. Similarly, deleting either of two families of active IS elements from the *Corynebacterium glutamicum* ATCC 13032 genome improved its properties [134]. The resulting WJ004 (with all four IS*Cg1* elements deleted) and WJ008 (with all four IS*Cg2* elements deleted) strains exhibited up to 6-fold increased electroporation efficiency and greater recombinant protein yields, while maintaining the same growth rates as the parental strain. Reduced-genome strains of *Bacillus subtilis* 168 can exhibit improved cell yield and enzyme or chemical production [62]. In this species, there is typically a gradual decrease in transformability as large regions of the chromosome are deleted [62,136]. However, this loss can be compensated for and transformation can even be improved over wild-type levels by overexpressing competence regulators [137]. Overall, our results with ADP1-ISx are broadly similar to studies of other clean-genome bacterial strains. We have explicitly shown that fewer inactivating mutations occur in ADP1-ISx, and it exhibits an unanticipated improvement in the efficiency of natural transformation.

CONCLUSIONS

Directed evolution is a powerful tool for creating novel biomolecules that do not exist in nature and for optimizing complex cellular systems. But, when inadvertent and unchecked, unplanned evolution can impede progress in bioengineering by introducing

mutations that lead to unpredictable failures of DNA-encoded functions. Our work constructing ADP1-ISx further highlights the effectiveness of a rational genome 'cleanup' strategy to improve a cellular chassis and the collateral benefits that often accompany domesticating an organism in this way. In microbial species that harbor a manageable number of IS elements, deleting these agents of instability may often be the simplest fail-safe approach for a high payoff in terms of improving genome stability and cell productivity. *Acinetobacter baylyi* ADP1-ISx is an improved chassis for synthetic biology and genetic studies that wish to take advantage of its natural transformability.

ACKNOWLEDGMENTS

This study was supported by the National Institutes of Health (R00-GM087550), Defense Advanced Research Projects Agency (HR0011-15-C0095), the National Science Foundation BEACON Center for the Study of Evolution in Action (DBI-0939454), and a National Science Foundation CAREER award (CBET-1554179).

SUPPLEMENTARY INFORMATION

Details of the mutations in ADP1-ISx-*rpoD-cyoB*, ADP1-ISx, and the 17 sequenced clones from the MA experiment are provided in **Table S1** (Appl. Env. Microbiol. 83: e01025-17. PMID: 28667117).

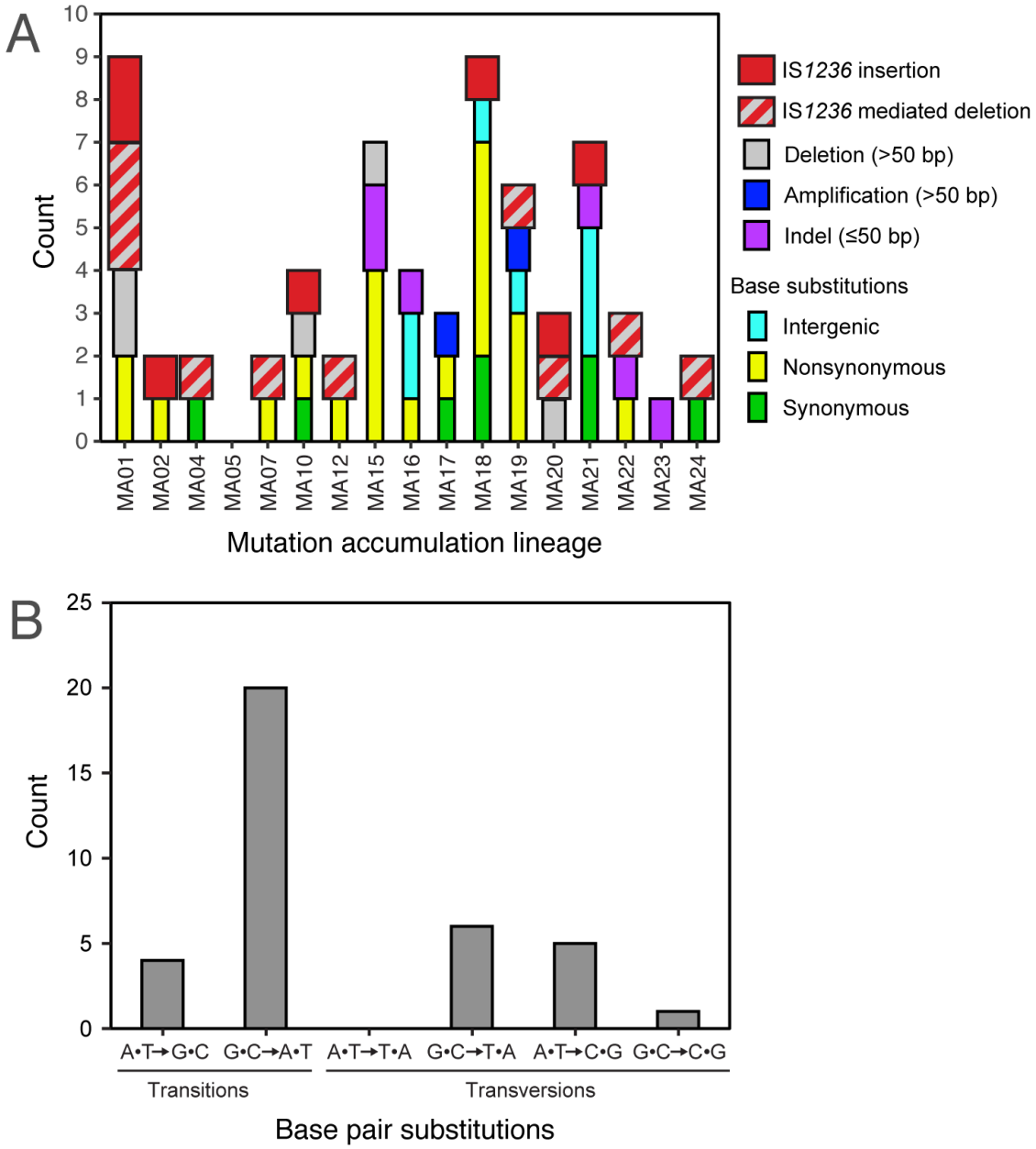

Figure 2.1. Mutations observed in *Acinetobacter bayli* ADP1 by whole-genome sequencing of 17 strains that each evolved separately for 7,500 generations under relaxed selection in the mutation accumulation (MA) experiment.

(A) Mutations observed in each clone from the MA experiment divided into major categories. Overall, 26% of mutations were directly related to IS*1236* element activity. The precise location and nature of each mutation is reported in **Table S1**. (B) Summed observations of each base pair substitution across all sequenced clones from the MA experiment.

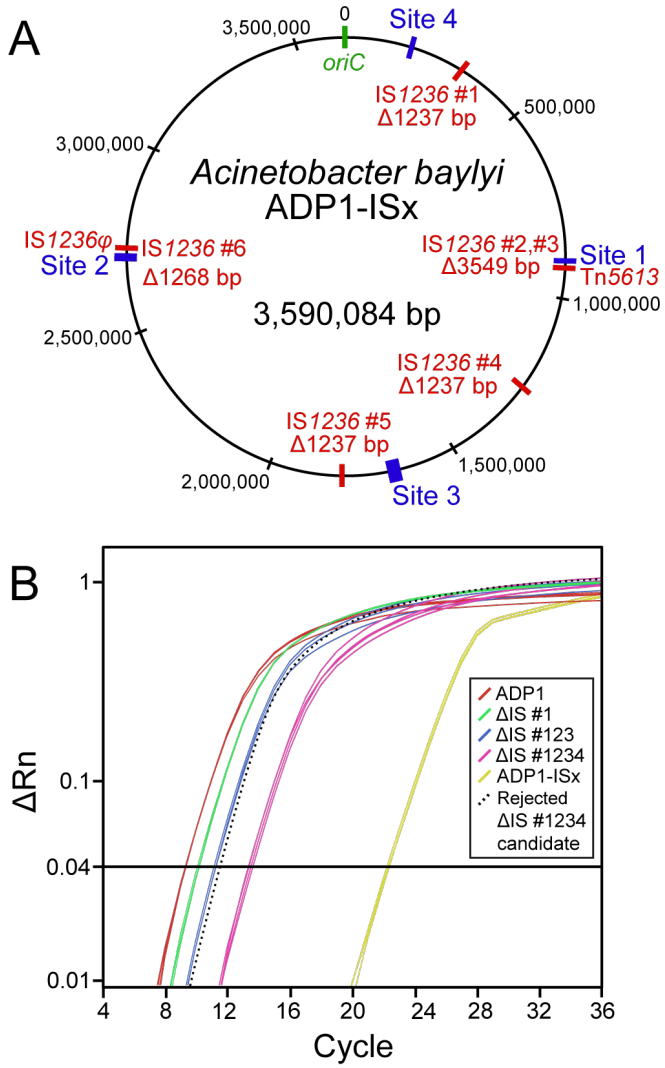

Figure 2.2. Construction of *Acinetobacter bayli* ADP1-ISx.

(A) Five unmarked genomic deletions were made to remove all six IS*1236* elements designated #1-6 found in wild-type ADP1 chromosome (shown in red) to create the ADP1-ISx strain. Sites 1-4 (shown in blue) were used for transformation and mutation assays. IS*1236* elements #2 and #3 form a composite transposon (Tn*5613*), and element #6 is inactive (IS*1236*φ). (B) Example of qPCR data used to monitor IS element deletion steps. The six IS copies per genome found in wild-type ADP1 register above a given fluorescence threshold (ΔRn value) during early PCR cycles when amplifying a 119-bp fragment located within the IS*1236* transposase gene. The sequential removal of IS elements in the deletion stains progressively increases the number of cycles necessary to reach this threshold from the same input quantity of genomic DNA. One example of a rejected candidate strain that accumulated a new IS element insertion elsewhere in the genome such that its IS copy number did not decrease after the deletion of IS #4 is shown (dashed line).

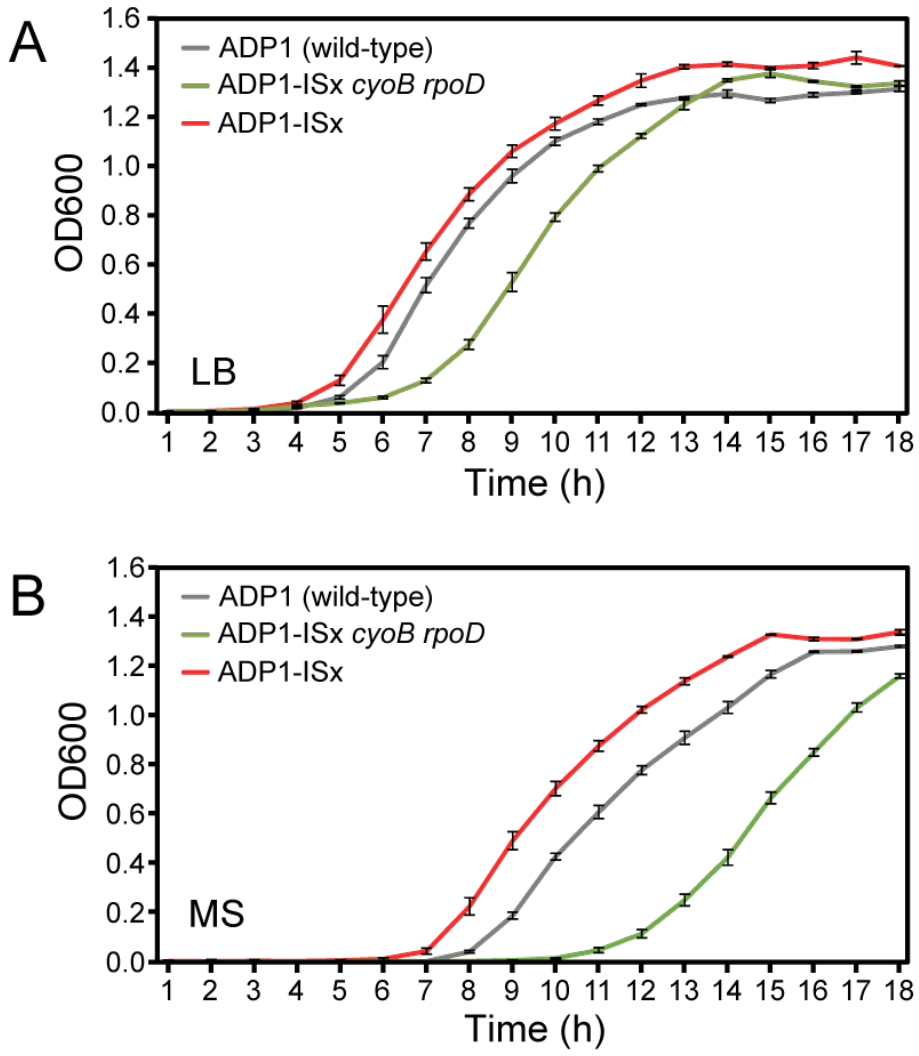

Figure 2.3. Repair of unintended *cyoB* and *rpoD* mutations sustained during ADP1-ISx strain construction.

Growth curves in LB broth (A) and MS defined medium (B) for wild-type ADP1, ADP1-ISx with the *cyoB* and *rpoD* mutations, and ADP1-ISx with these genes repaired to the wild-type sequence (as verified by whole-genome sequencing) were recorded by monitoring optical density at 600 nm (OD600). Error bars are standard deviations of three biological replicates.

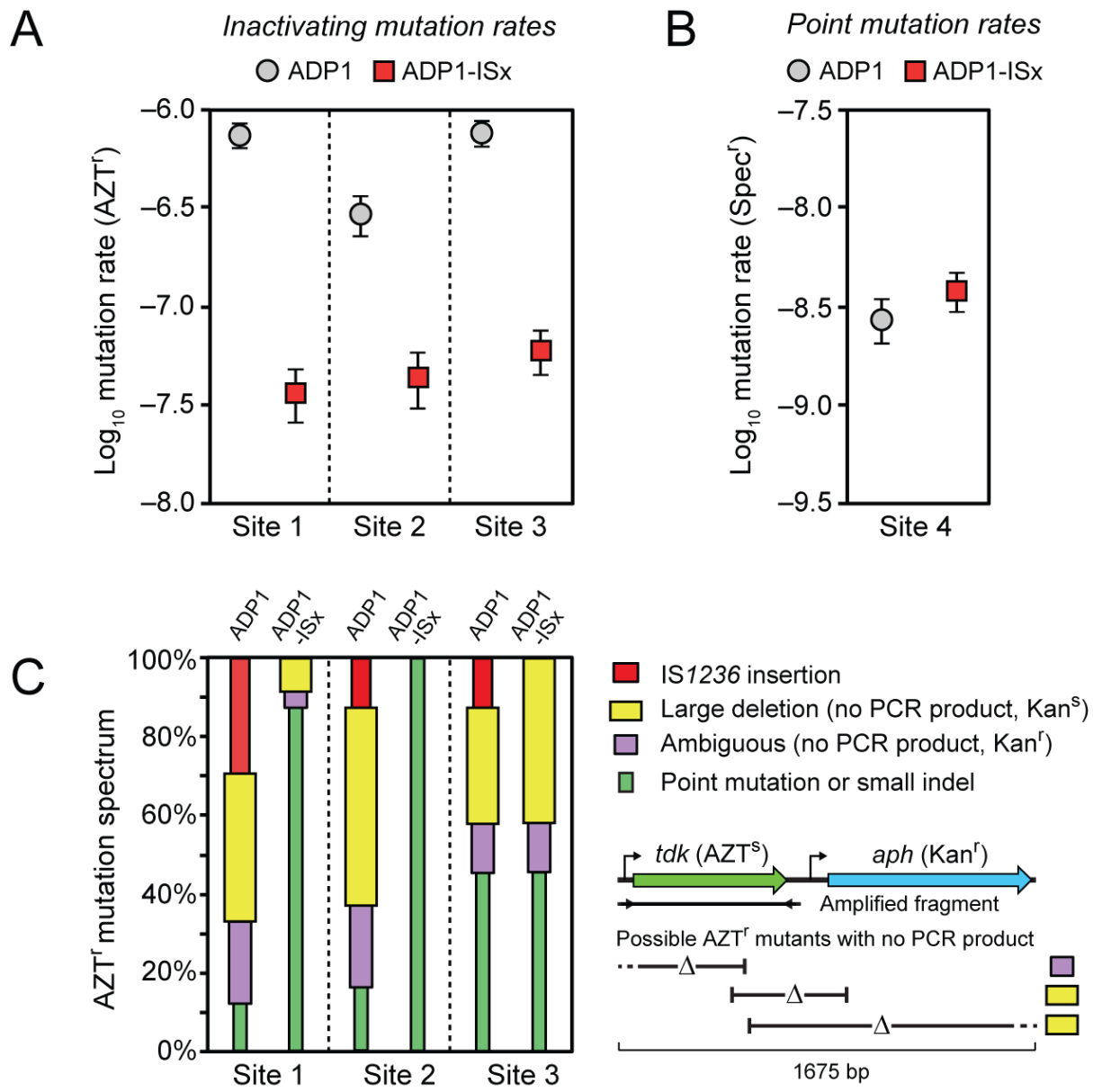
A
Inactivating mutation rates
ADP1
ADP1-ISx
Log10 mutation rate (AZTr)
–6.0
–6.5
–7.0
–7.5
–8.0
Site 1
Site 2
Site 3
B
Point mutation rates
ADP1
ADP1-ISx
Log10 mutation rate (Specr)
–7.5
–8.0
–8.5
–9.0
–9.5
Site 4
C
ADP1
ADP1 -ISx
AZTr mutation spectrum
100%
80%
60%
40%
20%
0%
IS1236 insertion
Large deletion (no PCR product, Kans)
Ambiguous (no PCR product, Kanr)
Point mutation or small indel
tdk (AZTs)
aph (Kanr)
Amplified fragment
Possible AZTr mutants with no PCR product
Δ
1675 bp

Figure 2.4

Figure 2.4. Reduced rates of inactivating mutations in ADP1-ISx.

(A) Rates of mutations leading to loss of function of a counterselectable marker gene (*tdk*) were determined using fluctuation tests that selected for resistance to azidothymidine (AZT^R). Error bars are 95% confidence intervals. The marker was placed at three different sites in the *A. baylyi* genome that differed in how near they were to IS*1236* elements in the wild-type ADP1 genome (see Fig. 2A). (B) Mutation rates were measured for reversion of a stop codon in the leader region of an antibiotic marker gene that restored spectinomycin resistance ($Spec^R$). This assay is expected to primarily reflect the point mutation rate. Error bars are 95% confidence intervals. (C) Spectrum of inactivating mutations determined by using PCR to amplify a fragment containing the *tdk* gene from the genome of one AZT^R resistant mutant from each of the 24 biological replicates of the fluctuation tests in part A. IS*1236* mutations and point mutations were inferred from an expected side change or no size change in this fragment, respectively. When no PCR product resulted and the function of the adjacent *aph* marker gene that was inserted alongside the *tdk* gene was also lost (yielding a Kan^S versus Kan^R phenotype), then the inactivating mutation was inferred to be a large deletion. If there was no PCR product and mutant remained Kan^R, then it was likely a large deletion that overlapped the *tdk* gene but not the *aph* gene. However, since the PCR reaction could have failed to amplify a band for other reasons, we conservatively classified this as an "ambiguous" result. This logic for classifying mutations is illustrated in the inset below the legend and further explained in the text (see *Altered mutational spectrum in ADP1-ISx*). As expected, no IS*1236* insertions were found in ADP1-ISx. The proportion of large deletions among the inactivating mutations was also greatly reduced at Site 1 and Site 2.

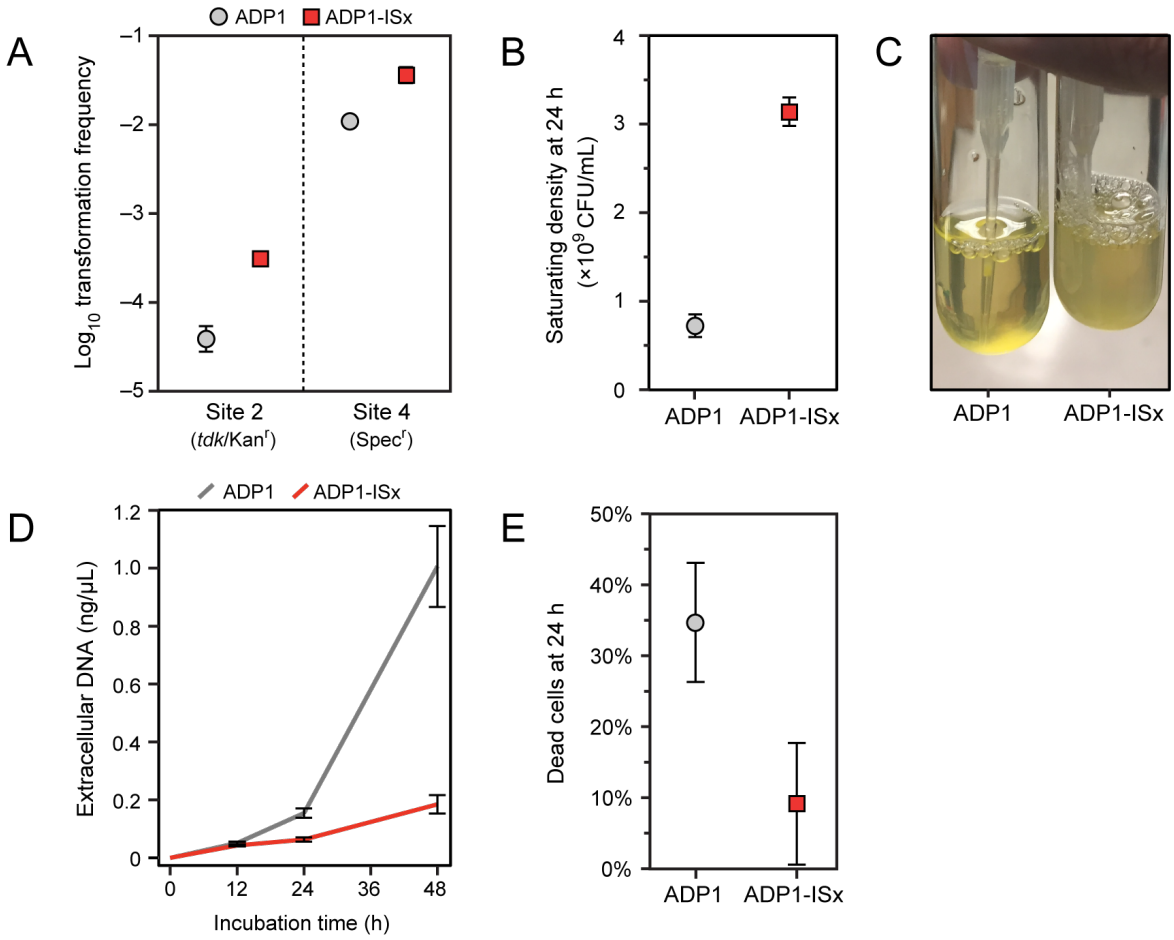

Figure 2.5. ADP1-ISx exhibits increased transformability and reduced autolysis.

(A) Transformation frequencies measured as transformants per colony-forming unit (CFU) under normal growth conditions in LB for PCR products containing the *tdk*-Kan^R cassette (into Site 2) or a spectinomycin resistance marker (into Site 4). (B) CFUs measured by plating dilutions of cultures on LB-agar after saturating growth. (C) Reduced settling behavior of ADP1-ISx compared to ADP1 in saturated cultures left at room temperature for 48 h after growth. (D) Accumulation of extracellular DNA over time in the supernatant of MS cultures grown under standard conditions. (E) Fraction of cells staining as dead in LB cultures after 24 hours of growth.

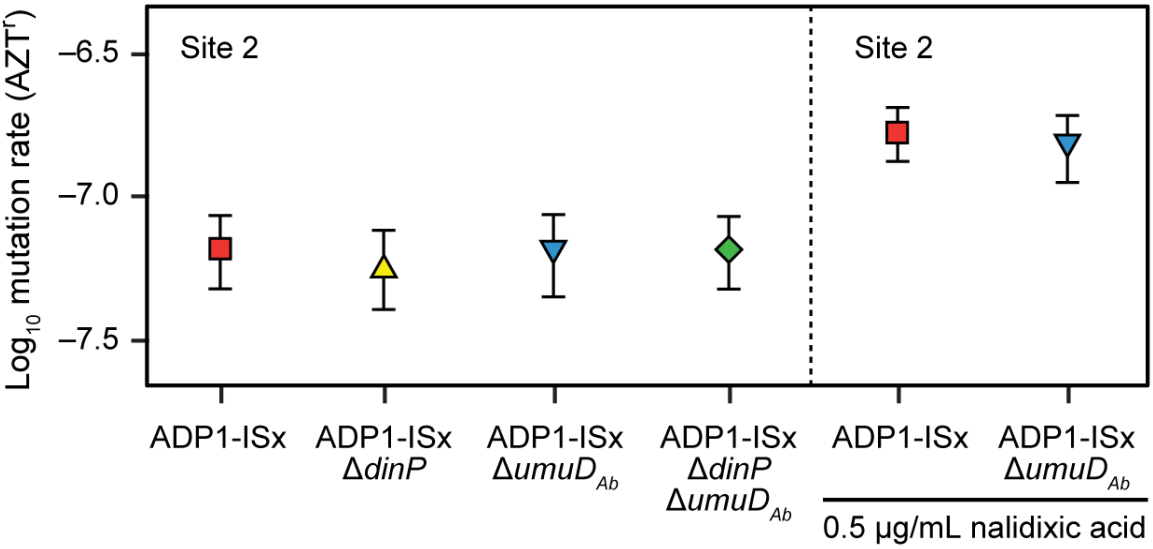

Figure 2.6. Inactivating mutation rates are unchanged in ADP1-ISx after deletion of error prone polymerase DinP and DNA damage response regulator UmuDAb.

The effects of deleting each gene on the rates of mutations inactivating the *tdk* counterselectable marker gene inserted into the genome at Site 2 were measured for each strain. Mutation rates in ADP1-ISx and the *umuDAb* deletion strain were also determined in the presence of a subinhibitory concentration of nalidixic acid, which induces double-strand breaks in DNA. Error bars are 95% confidence intervals.

Primer	Description	Sequence
***tdk*/Kan^R cassette**		
P1-F	*tdk*/Kan^R amplification for deletion and rescue cassettes	CCCAGCCTCCAATTCAAATCA
P2-R	*tdk*/Kan^R amplification for deletion and rescue cassettes	CCAGCTCCGCATGCTTAGAA
tdk-R	Used with P1-F to amplify just the *tdk* gene from the *tdk*/KanR cassette	ACATCAGAGATTTTGAGACACAACGTG
Deletion of IS*1236* elements to create ISx-ADP1		
IS#1-P3-F	Amplify 5′ flank for IS*1236* #1 deletion and rescue cassettes	GCTTGAACCTCACTCCTGA
IS#1-P1′P4-R	Amplify 5′ flank with IS#1-P3-F for IS*1236* #1 deletion cassette	TGATTTGAATTGGAGGCTGGGACAATT GCATAAAAATGAAAGTAACTC
IS#1-P2′P5-F	Amplify 3′ flank with IS#1-P6-R for IS*1236* #1 deletion cassette	TTCTAAGCATGCGGAGCTGGTGTCCAG TCAATTGCTTTAATG
IS#1-P6-R	Amplify 3′ flank for IS*1236* #1 deletion and rescue cassettes	TCGATAGACTAAAATGACTAGACAGG
IS#1-P5′P4-R	Amplify 5′ flank with IS#1-P3-F for IS*1236* #1 deletion cassette	CATTAAAGCAATTGACTGGACAACAAT TGCATAAAAATGAAAGTAACTC
IS#1-P4′P5-F	Amplify 3′ flank with IS#1-P6-R for IS*1236* #1 deletion cassette	GAGTTACTTTCATTTTTATGCAATTGTT GTCCAGTCAATTGCTTTAATG
IS#23-P3-F	Amplify 5′ flank for Tn*5613* deletion and rescue cassettes	TAAAGGACCTGATAAGGCGAT
IS#23-P1′P4-R	Amplify 5′ flank with IS#23-P3-F for Tn*5613* deletion cassette	TGATTTGAATTGGAGGCTGGGCTTTAA TGATTAAATAATCAATTTTCATCT
IS#23-P2′P5-F	Amplify 3′ flank with IS#23-P6-R for Tn*5613* deletion cassette	TTCTAAGCATGCGGAGCTGGTCAATTG ACCGCTTGATCA
IS#23-P6-R	Amplify 3′ flank for Tn*5613* deletion and rescue cassettes	GATCTATTTGGCAATATCATAAACATC
IS#23-P5′P4-R	Amplify 5′ flank with IS#23-P3-F for Tn*5613* deletion cassette	TGATCAAGCGGTCAATTGACTTTAATG ATTAAATAATCAATTTTCATCT
IS#23-P4′P5-F	Amplify 3′ flank with IS#23-P6-R for *Tn5613* deletion cassette	AGATGAAAATTGATTATTTAATCATTA AAGTCAATTGACCGCTTGATCA
IS#4-P3-F	Amplify 5′ flank for IS*1236* #4 deletion and rescue cassettes	GGCGATTAAGAGAGTAATTCCA
IS#4-P1′P4-R	Amplify 5′ flank with IS#4-P3-F for IS*1236* #4 deletion cassette	TGATTTGAATTGGAGGCTGGGATGAAC ATCTTCAACATTTAGATCAAG
IS#4-P2′P5-F	Amplify 3′ flank with IS#4-P6-R for IS*1236* #4 deletion cassette	TTCTAAGCATGCGGAGCTGGCATGCAT TGCACGTTCG
IS#4-P6-R	Amplify 3′ flank for IS*1236* #4 deletion and rescue cassettes	AATGCCAGATCTACACTGACG
IS#4-P5′P4-R	Amplify 5′ flank with IS#4-P3-F for IS*1236* #4 deletion cassette	CGAACGTGCAATGCATGATGAACATCT TCAACATTTAGATCAAG
IS#4-P4′P5-F	Amplify 3′ flank with IS#4-P6-R for IS*1236* #4 deletion cassette	CTTGATCTAAATGTTGAAGATGTTCATC ATGCATTGCACGTTCG
IS#5-P3-F	Amplify 5′ flank for IS*1236* #5 deletion and rescue cassettes	GAATTTAAAGTTAAAGATATTTTTGCG

Table 2.1. Primers used in this study

Primer	Description	Sequence
IS#5-P1′P4-R	Amplify 5′ flank with IS#5-P3-F for IS*1236* #5 deletion cassette	TGATTTGAATTGGAGGCTGGGGTTTGA GTTTATCGGAAAAACTATC
IS#5-P2′P5-F	Amplify 3′ flank with IS#5-P6-R for IS*1236* #5 deletion cassette	TTCTAAGCATGCGGAGCTGGAACAAAA GGCCATGGAG
IS#5-P6-R	Amplify 3′ flank for IS*1236* #5 deletion and rescue cassettes	GTACCTTGCCGAGCAAA
IS#5-P5′P4-R	Amplify 5′ flank with IS#5-P3-F for IS*1236* #5 deletion cassette	CTCCATGGCCTTTTGTTGTTTGAGTTTAT CGGAAAAACTATC
IS#5-P4′P5-F	Amplify 3′ flank with IS#5-P6-R for IS*1236* #5 deletion cassette	GATAGTTTTTCCGATAAACTCAAACAAC AAAAGGCCATGGAG
IS#6-P3-F	Amplify 5′ flank for IS*1236* #6 deletion and rescue cassettes	GGAGTATCGTGATCCCAAAG
IS#6-P1′P4-R	Amplify 5′ flank with IS#6-P3-F for IS*1236* #6 deletion cassette	TGATTTGAATTGGAGGCTGGGCTACTC AGTTAGTCGAAAGTATCATCC
IS#6-P2′P5-F	Amplify 3′ flank with IS#6-P6-R for IS*1236* #6 deletion cassette	TTCTAAGCATGCGGAGCTGGGTTCAAG TGGGGTTTACTGG
IS#6-P6-R	Amplify 3′ flank for IS*1236* #6 deletion and rescue cassettes	CAGTGATATGCATGCTGAGG
IS#6-P5'P4-R	Amplify 5′ flank with IS#6-P3-F for IS*1236* #6 deletion cassette	CCAGTAAACCCCACTTGAACCTACTCA GTTAGTCGAAAGTATCATCC
IS#6-P4'P5-F	Amplify 3′ flank with IS#6-P6-R for IS*1236* #6 deletion cassette	GGATGATACTTTCGACTAACTGAGTAG GTTCAAGTGGGGTTTACTGG
qPCR to monitor IS*1236* deletion		
qIS-F	Amplify 119-bp fragment common to all IS*1236* elements with qIS-R	GTTTTCGCCAGGCATAATA
qIS-R	Amplify 119-bp fragment common to all IS*1236* elements with qIS-F	CAGATCATGCCAAGAAAAGTAC
Repair of *rpoD* and *cyoB* mutations		
rpoD-F	Amplify fragment containing *rpoD* mutation with *rpoD*-R	GCCTTCAAACAATTCACTTAACC
rpoD-R	Amplify fragment containing *rpoD* mutation with *rpoD*-F	GCAATCGAAATAACGAGACG
rpoD-S	Sequence mutation in *rpoD* gene	GCTGAGGTTAACGATCATCT
cyoB-F	Amplify fragment containing *cyoB* mutation with *cyoB*-R	GGTACGCTTTTATCTGGTGTAA
cyoB-R	Amplify fragment containing *cyoB* mutation with *cyoB*-F	CAGCAGAATTATGATCATGACTC
cyoB-S	Sequence mutation in *cyoB* gene	GACACGTCGTTTGAATACAT
Integration of *tdk*/Kan^R cassette for measuring inactivating mutation rates and transformation frequency		
Site1-P3-F	Amplify 5′ flank for integration of *tdk*/Kan^R cassette	ATAAAAAATTTTTTATTATAATTAATA ATTAAGTTGTG
Site1-P1′P4-R	Amplify 5′ flank with Site1-P3-F for integration of *tdk*/Kan^R cassette	TGATTTGAATTGGAGGCTGGGATCTAA ACCTTGTGATATCAACTGTT
Site1-P2′P5-F	Amplify 3′ flank with Site1-P6-R for integration of *tdk*/Kan^R cassette	TTCTAAGCATGCGGAGCTGGGTGCAAA GAATATTCATCTGAATG

Table 2.1. Primers used in this study (continued)

Primer	Description	Sequence
Site1-P6-R	Amplify 3′ flank for integration of *tdk*/Kan^R cassette	TACTTCAGATTCTGGGAATTTCT
Site2-P3-F	Amplify 5′ flank for integration of *tdk*/Kan^R cassette	CAAATCCATTGAAAAGAAATGC
Site2-P1′P4-R	Amplify 5′ flank with Site2-P3-F for integration of *tdk*/Kan^R cassette	TGATTTGAATTGGAGGCTGGGCTGATC ATCATGGACAAGCT
Site2-P2′P5-F	Amplify 3′ flank with Site2-P6-R for integration of *tdk*/Kan^R cassette	TTCTAAGCATGCGGAGCTGGTGTGATC TTTTTTTCAAATTCATT
Site2-P6-R	Amplify 3′ flank for integration of *tdk*/Kan^R cassette	GGTAATGTTCTCTCCTACGAAAAG
Site3-P3-F	Amplify 5′ flank for integration of *tdk*/Kan^R cassette	CCTGATCCAGAAAGATATATGACTT
Site3-P1′P4-R	Amplify 5′ flank with Site3-P3-F for integration of *tdk*/Kan^R cassette	TGATTTGAATTGGAGGCTGGGTGAAAA GTTATCAAATATTGACGTATG
Site3-P2′P5-F	Amplify 3′ flank with Site3-P6-R for integration of *tdk*/Kan^R cassette	TTCTAAGCATGCGGAGCTGGTAAAGTG ATCGAGCAGCCA
Site3-P6-R	Amplify 3′ flank for integration of *tdk*/Kan^R cassette	AACTATACATATTGAATTTTTAAATAAT AATTTAATTC
Integration of Spec^R genes in Region 1 for measuring point mutation rates and transformation frequency		
R1-P3-F	Amplify R1 5′ flank for integration of *tdk*/Kan^R cassette	ACGCCGAGTCCTCTTGAGTACAGG
R1-P1′P4-R	Amplify R1 5′ flank with Site1-P3-F for integration of *tdk*/Kan^R cassette	TGATTTGAATTGGAGGCTGGGGATTTT CCGCCCATCTCAC
R1-P2′P5-F	Amplify R1 3′ flank with Site1-P6-R for integration of *tdk*/Kan^R cassette	TTCTAAGCATGCGGAGCTGGCAGAAAT TATAAAACGCACATCA
R1-P6-R	Amplify R1 3′ flank for integration of *tdk*/Kan^R cassette	ACTCGCTGCAATAGTGGCAAAAGC
Spec^R-F	Amplify Spec^R gene to construct rescue cassette	TTCAAATATGTATCCGCTCATGAGACA ATAAC
Spec^R-R	Amplify Spec^R gene to construct rescue cassette	TTATTTGCCGACTACCTTGGTGATC
Spec^R-E8*-R	Amplify 5′ end of Spec^R gene with Spec^R-F to introduce stop codon	CTGATAGTTGAGTCGATACTTAGGCGA
Spec^R-E8*-F	Amplify 3′ end of Spec^R gene with Spec^R-R to introduce stop codon	GCCTAAGTATCGACTCAACTATCAGAG GTA
spec^R-R1-R	Amplify R1 5′ flank with R1-P3-F to create overlap with Spec^R gene	ATGAGCGGATACATATTTGAAGATTTT CCGCCCATCTCAC
spec^R-R1-F	Amplify R1 3′ flank with R1-P6-R to create overlap with Spec^R gene	ACCAAGGTAGTCGGCAAATAATCTTTA CGAGTCATGCCAGCAC
Deletion of *dinP* and *umuDAb*		
dinP-P3-F	Amplify 5′ flank for *dinP* deletion and rescue cassettes	CAAAGCGATAATGTAGAAAAAAC
dinP-P1′P4-R	Amplify 5′ flank with *dinP*-P3-F for *dinP* deletion cassette	TGATTTGAATTGGAGGCTGGGTCGTA TTTTTTTTACACAAAAA

Table 2.1. Primers used in this study (continued)

Primer	Description	Sequence
dinP-P2′P5-F	Amplify 3′ flank with *dinP*-P6-R for *dinP* deletion cassette	TTCTAAGCATGCGGAGCTGGTCATTTT ACTGTATTTGTGATTT
dinP-P6-R	Amplify 3′ flank for IS*1236* #1 deletion and rescue cassettes	ATGCTCTCGATGAGTATTGG
dinP-P5′P4-R	Amplify 5′ flank with *dinP*-P3-F for *dinP* deletion cassette	CTCAAATCACAAATACAGTAAAATGA TCGTATTTTTTTTACACAAAAATTTAA
dinP-P4′P5-F	Amplify 3′ flank with *dinP*-P6-R for *dinP* deletion cassette	TTAAATTTTTGTGTAAAAAAAATACG ATCATTTTACTGTATTTGTGATTTGAG
umuD-P3-F	Amplify 5′ flank for *umuD* deletion and rescue cassettes	TTGGCTCCACTACTCACAGA
umuD-P1′P4-R	Amplify 5′ flank with *umuD*-P3-F for *umuD* deletion cassette	TGATTTGAATTGGAGGCTGGGTCATG AGTCAGAGAATCTTTGC
umuD-P2′P5-F	Amplify 3′ flank with *umuD*-P6-R for *umuD* deletion cassette	TTCTAAGCATGCGGAGCTGGAATGTT TTCTCAAGTTAAAATAATCTAA
umuD-P6-R	Amplify 3′ flank for *umuD* deletion and rescue cassettes	ATCTATACTAGTAGATTATACGGACGA TG
umuD-P5′P4-R	Amplify 5′ flank with *umuD*-P3-F for *umuD* deletion cassette	TTAGATTATTTTAACTTGAGAAAACAT TTCATGAGTCAGAGAATCTTTGC
umuD-P4′P5-F	Amplify 3′ flank with *umuD*-P6-R for *umuD* deletion cassette	GCAAAGATTCTCTGACTCATGAAATGT TTTCTCAAGTTAAAATAATCTAA

Table 2.1. Primers used in this study (continued)

Chapter 3: Rapid and assured genetic engineering methods applied to *Acinetobacter baylyi* ADP1 genome streamlining

Suárez, G.A., Dugan, K.R., Renda B.A.[‡], Leonard S.P., Gangavarapu L.S., Barrick, J.E.

Research Contributions

Suárez, G.A. designed the experiments and methodology. Suárez, G.A., Dugan, K.R. performed all CRISPR experiments. Renda B.A. and Leonard S.P. performed and analyzed Tn-seq experiments. Gangavarapu L.S. contributed with DNA sample processing and validation. Suárez, G.A., and Barrick J.E. analyzed data. Suárez, G.A. and Barrick, J.E. wrote the manuscript.

Acknowledgements

We also thank Daniel Deatherage for sequencing pipeline tips and technical support. We acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for access to high-performance computing resources.

ABSTRACT

Precision engineering of bacterial genomes requires assembling DNA modules and integrating them at defined chromosomal sites without introducing off-target mutations. One goal of synthetic biology is to improve the efficiency and predictability of living cells by removing extraneous genes from their genomes. Here, we demonstrate improved methods for engineering the genome of the metabolically versatile and naturally transformable bacterium *Acinetobacter baylyi* ADP1. In Golden Transformation, linear DNA fragments constructed by Golden Gate Assembly are directly added to cells to create targeted deletions, edits, or additions to the chromosome. We tested the dispensability of 55 large regions of the ADP1 chromosome using Golden Transformation. The 19 successful multiple-gene-deletions ranged in size from 21 to 118 kilobases and collectively accounted for 24.6% of its genome. Deletion success could not be predicted on the basis of essentiality predictions from a single-gene knockout collection or a Tn-seq experiment we conducted in rich media. We further show that the native CRISPR/Cas locus is active in ADP1 and that it can be retargeted using Golden Transformation. We reprogrammed it to demonstrate that a self-targeting CRISPR-lock can be used to validate that genes have been successfully removed from the chromosome and to prevent them from being reacquired. These methods can be used together to implement combinatorial routes to further genome streamlining and for more rapid and assured metabolic engineering of this versatile chassis organism.

INTRODUCTION

Biological engineering faces unique challenges that must be addressed before it will be as rapid, reliable, and robust as traditional engineering disciplines [45,138]. On one front, simpler and more flexible genetic engineering tools can facilitate re-writing the genome sequences of living cells to speed up the design-build-test cycle. Another major challenge is the complexity inherent in using living cells as a starting point. Even 'simple' microbial cells maintain thousands of distinct components, from small molecules to proteins, but only a smaller subset of these are absolutely required for cellular replication. In fact, many of these functions are unnecessary or even undesirable in controlled laboratory and industrial settings that are not as varied as the natural challenges that shaped the evolution of these cells. Therefore, a major aim of the field of synthetic biology has been to re-factor and re-engineer genomes to create 'chassis' organisms with reduced complexity [139,140]. The main approaches to genome simplification fall into two categories: a bottom-up approach in which DNA pieces are assembled into a minimal genome that is tested to see if it can successfully 'boot up' self-replication in a host cell and a top-down approach in which chromosomal regions found to be dispensable are removed in a step-wise fashion to create a streamlined genome. These efforts have led to more efficient and reliable microbial cell factories and blurred the line between chemical and living systems [141–144].

Bacterial genome streamlining efforts often begin by deleting regions of a chromosome that encode selfish or parasitic sequences that are obviously unnecessary

and may even be hazardous to a cell. Deleting transposons and prophage sequences reduces mutation rates and improves the fitness of cells [52,68,124,137,138]. It is also relatively straightforward to recognize and then delete gene clusters encoding resource- or energy-intensive functions that cells perform to survive in the wild but are not needed in laboratory or industrial cultures. For example, bacterial growth and productivity has been improved by deleting motility systems, such as flagella [148], and large genomic islands encoding pathways for synthesizing secondary metabolites [67,149].

For more extreme genome reduction, information about the conservation of genes within a bacterial species and what single-gene knockouts are viable can be integrated to predict additional genome regions to target for deletion [54,55,68]. Computational tools are emerging that further integrate gene essentiality predictions with whole-cell metabolic and gene expression models to predict the viability of strains that combine multiple deletions [88]. Still, viable routes to extreme genome reduction remain difficult to predict because there can be interactions between genes that make some combinations of deletions inviable (e.g., synthetic lethality). Even when multiple large chromosomal regions can be removed, significant growth defects often result from accumulating losses of 'quasi-essential' genes, defined as genes that are not needed for survival but are required for robust growth [79–84]. Given how difficult it is to predict how gene essentiality and quasi-essentiality generalizes across environments [150], it is often necessary to empirically test the viability and performance of many different deletions and combinations of those deletions to create a useful minimal genome organism.

Acinetobacter baylyi ADP1 is a non-pathogenic soil bacterium that is notable for its naturally transformability under normal laboratory growth conditions [1,2]. ADP1 has been used in fundamental studies of DNA uptake and mechanisms of bacterial genome evolution [16,151,152]. Multiple studies have also highlighted its useful innate metabolic abilities, particularly related to degradation of aromatic compounds [153–157] and production of triacylglycerols and wax esters [30,158]. ADP1's natural transformability has been exploited for targeted mutagenesis and disruption of genes in its chromosome for metabolic engineering of these pathways [38,159], as well as to add novel genes to its genome and evolve them to improve these functions [30,158,160]. The natural transformability of ADP1 makes it an especially tractable system for testing genome streamlining approaches and their impact on engineered functions.

We previously began the process of streamlining the genome of *A. baylyi* ADP1 by creating a transposon-free strain that exhibited lower mutation rates, enhanced transformation, and reduced autolysis [146]. Here, we generated a new Tn-Seq dataset defining gene essentiality in rich medium and used this information to design 55 multiple-gene-deletions covering 58.3% of its chromosome. To facilitate engineering these large deletions, we developed a "Golden Transformation" genome engineering procedure that uses Golden Gate Assembly to construct the necessary DNA modules for transformation [161,162]. Nineteen of these large deletions, together comprising 24.6% of the ADP1 genome, were viable. Golden Transformation can also be used to facilitate other genome engineering tasks such as constructing and inserting gene expression modules into the chromosome sequence modules. Finally, we demonstrate that the native

CRISPR/Cas locus in *A. baylyi* ADP1 is active and show that it can be reprogrammed to block reacquisition of deleted genomic regions. Overall, our work demonstrates how the native genetic capabilities of *Acinetobacter baylyi* ADP1 make it an especially tractable system for testing combinatorial routes to more extreme genome streamlining and engineering.

MATERIALS AND METHODS

Culture conditions

Acinetobacter baylyi ADP1 [23] and ADP1-ISx [59] were grown at 30°C in LB-Miller (10 g NaCl, 10 g tryptone, 5 g yeast extract per liter). Liquid cultures were incubated in 8×150mm test tubes with orbital shaking at 200 r.p.m. over a 1-inch diameter. Solid media contained 1.6% agar (w/v). Media were supplemented with 50 μg/mL kanamycin (Kan), 60 μg/mL spectinomycin (Spec), 20 μg/mL gentamycin (Gent), 200 μg/mL 3-azido-2′,3′-dideoxythymidine (AZT), and 0.3 mM diaminopimelic acid (DAP), as indicated.

Golden Gate assembly

DNA fragments for transformation were constructed using Golden Gate Assembly (GGA) reactions containing 1 U/μl BsaI-HF or 0.5 U/μl BsmBI and 150 U/μl of either T7 or T4 DNA ligase in T4 DNA ligase buffer (New England Biolabs) [163,164]. The standard ~5-hour protocol was 31 cycles, each consisting of 37°C at 5 min and 16°C at 5 min, followed by one-time final incubations at 55°C for 5 min and 80°C for 5 min.

For the short ~1-hour protocol, the 37°C and 16°C steps were reduced to 1 min apiece. The pBTK622 part plasmid was constructed by adding the *tdk-kanR* cassette [10] to the pYTK001 entry vector as described elsewhere [164]. Plasmid samples input into GGA reactions were isolated from *E. coli* cultures using PureLink Quick Plasmid Miniprep Kit (ThermoFisher). Input PCR products were purified using the GeneJet PCR purification kit (ThermoFisher). Concentrations of all DNA parts were determined using the Qubit dsDNA BR Assay Kit (Invitrogen).

ADP1-ISx transformation

ADP1-ISx glycerol stocks were streaked out on LB agar and incubated overnight. For each transformation assay replicate, a different colony was picked and inoculated onto LB and grown overnight to produce a culture ready for transformation. Transformations in liquid culture were initiated by mixing input DNA, 250 µL LB, and 17.5 µL of the overnight ADP1-ISx culture. Puddle transformations were initiated by mixing input DNA with 50 µL of the overnight ADP1-ISx culture and transferring the complete volume onto a 13 mm diameter, 0.025 µm pore size mixed cellulose esters membrane (Millipore #VSWP01300) that was then placed on the surface of an LB agar plate. In both cases, transformations were incubated for 24 h under normal *A. baylyi* growth conditions. Then, dilutions in sterile saline were plated on LB and LB-Kan agar to estimate transformation frequencies or on LB-Kan or LB-AZT agar for genome modification steps.

Transformation frequency assays

Overlap extension PCR (OE-PCR) transformations used 0.01 pmol of a 3,875-bp DNA product constructed by joining the *tdk-kanR* cassette to 1.2 kb of 5' homology and 1.0 kb of 3'-homology to "Site 2" in the ADP1 genome as previously described [59]. GGA DNA transformations used 10 μL assembly reactions that included 0.01 pmol of each DNA part: the pBTK622 plasmid and PCR amplicons that added the necessary restriction sites to the same two flanking genomic homology regions. Multiple GGA reactions were pooled and re-divided when testing the same DNA construct in multiple replicates and across conditions. Primer sequences for creating the PCR products used in each type of assembly are provided in **Table 3.1**.

Transformation frequencies were determined by comparing colony-forming units (CFUs) on LB-Kan agar to CFUs on LB agar with 3-4 replicates for each condition. Puddle transformations were resuspended by carefully transferring the filter with sterile tweezers into 10 mL of saline and vortexing to release cells before plating dilutions. Negative controls without DNA yielded no colonies on LB-Kan agar. Transformation frequencies and confidence intervals were estimates from CFU values using Poisson regression models in R [165].

Tn-Seq

The transposon mutagenized library of *A. baylyi* ADP1 was constructed through conjugation of the mini-transposon suicide vector pBT20 [166]. For the conjugation, overnight cultures of *E. coli* β2163 [167] carrying pBT20 and ADP1 were pelleted via centrifugation and resuspended together at a 3:1 ratio in 100 μL of sterile saline. This

mixture was then spotted onto the same type of membrane used in puddle transformations and placed on an LB-DAP plate. After ~24 hours at 30°C, the filter was transferred to a 1.7 mL Eppendorf tube with 1 mL of sterile saline. After vortexing to resuspend cells, 800 μL of this mixture was used to inoculate 50 mL of LB-Gent. After overnight growth, 1.2 mL aliquots of this culture were stored at −80°C as glycerol stocks.

For sequencing library preparation, genomic DNA was extracted from the frozen aliquots using a Purelink Genomic DNA Mini Kit (ThermoFisher) and sheared on a Covaris S220 to an average size of 300 bp. Poly-C tails were added to the fragmented DNA using terminal deoxynucleotidyl transferase (Promega). Then, the processed fragments were size-selected using AMPure XP beads (Beckman Coulter) and amplified by PCR with a biotinylated primer binding inside of the transposon and a non-biotinylated primer binding to the poly-C tail. The amplified library fragments were then bound to streptatividin M-280 Dynabeads (ThermoFisher) and washed to remove unbound DNA. Illumina barcodes and adaptors were added in a subsequent PCR reaction in which internal barcodes were introduced to allow multiple libraries to be run with the same external Illumina barcode. The libraries were pooled in equal ratios, analyzed on a BioAnalyzer 2100 (Agilent) for fragment distribution, and sequenced on an Illumina HiSeq 4000 at the University of Texas at Austin Genome Analysis and Sequencing Facility (GSAF). Tn-Seq FASTQ files are available from the NCBI Sequence Read Archive (PRJNA559727).

Tn-seq reads were analyzed as previously described [168]. Briefly, reads containing transposon sequences were quality filtered, trimmed of adapter sequences, and then

mapped to the *A. baylyi* ADP1 genome (GenBank: NC_005966.1) [23]. To identify essential genes, we compared the observed number of insertion mutants in each gene to simulated null distributions based on the total number of possible insertion sites. Genes designated as "essential" were significantly depleted for transposon mutants when compared to the simulated null distributions using DESeq2 [169]. Scripts for performing this analysis are freely available online (https://github.com/spleonard1/Tn-seq). The functional categories of genes in the *A. baylyi* ADP1 genome were predicted using eggNOG-mapper (version 4.5.1) [170]. The full output of the Tn-Seq analysis with COG assignments for each gene is shown in **Table S1**.

Multiple-gene-deletion Strain Construction

Large deletion variants of ADP1-ISx were generated by transforming GGA reactions, with the exception of MGD7 and MGD11, which were generated by transforming OE-PCR reactions. GGA reactions and transformations were performed as described above except we used 40 µL GGA reactions containing 500 ng of pBTK622 and 600 ng of each of the ~2-kb flanking genomic homology fragments generated by PCR. When standard transformations of a GGA reaction for a deletion yielded no colonies or very few colonies on LB-Kan agar, we attempted puddle transformation of a new GGA reaction.

Genome Sequencing

Genomic DNA isolated from candidate large-scale deletion strains was prepared for sequencing as described previously [171]. Briefly, genomic DNA was extracted from

overnight cultures, then fragmented, end-repaired, A-tailed, ligated to adapters, and size-selected for library preparation. The resulting DNA libraries were sequenced to >100× coverage on an Illumina HiSeq 2500 platform at the UT Austin GSAF. Read files were analyzed with *breseq* (version 0.28.0) [172,173] to predict mutations, large deletions and other types of structural variation, relative to the *A. baylyi* ADP1 genome. FASTQ files from genome sequencing are available from the NCBI Sequence Read Archive (PRJNA559727).

Growth rates

The growth rates of 'unrescued' multiple-gene-deletion strains that still contained the tdk-kanR cassette in their genomes were compared to growth rates of ADP1-ISx. For each strain being tested, 2 μL of a glycerol stock was inoculated into 200 μL of LB in a 96-well clear flat-bottom microplate (Costar). This microplate was incubated overnight at 30°C with 250 r.p.m. orbital shaking over a diameter of 1-inch. Subsequently, 2 μL of the preconditioned well for each strain was used to inoculate 200 μL of fresh LB in each of three replicate wells in a new microplate. The optical density at 600 nm (OD600) was recorded every 10 minutes following a brief shaking step (15 seconds with a 6 mm amplitude) as these cultures were incubated at 30°C in an Infinite M200 Pro microplate reader (Tecan).

OD600 values were corrected by subtracting measurements of blank wells and then further normalized by offsetting all measurements in each well such that the first three points had the same average as the grand mean of these points across all inoculated

wells on the entire plate. Growth rates were determined from nonlinear least squares fits to an exponential growth equation with lag and doubling time parameters in R [165]. The maximum specific growth rate fit from any set of 7 consecutive points (spanning 1 h of measurements) that all had OD600 values between 0.01 and 0.15 was determined for each well.

CRISPR assays

The 'CRISPR-ready' variant of ADP1-ISx was created by replacing all but one of the 91 spacers of its native CRISPR array with the *tdk-kanR* cassette using BsaI Golden Transformation and selection on LB-Kan. CRISPR reprogramming was accomplished by transforming the CRISPR-ready strain with a rescue cassette encoding a new spacer sequence. To assemble each spacer cassette, two complementary oligonucleotides were synthesized and annealed to create a double-stranded insert flanked by restriction sites. The spacer was then added to the genome via BsmBI Golden Transformation using the same two homology flanks used in the replacement step and selection on LB-AZT agar. Proper insertion of each spacer into the ADP1-ISx genome was verified by PCR and Sanger sequencing. Transformation assays for acquisition of spectinomycin resistance were conducted as described above but using 100 ng of donor DNA from an ADP1 strain constructed by replacing bases 139832–140099 of the genome (GenBank: NC_005966.1), which overlap the nonessential gene *ACIAD0135*, by transforming an OE-PCR product containing the *specR* cassette from pIM1463 and flanking homology [117].

In the CRISPR-Lock experiment, CRISPR-ready variants of the MGD2 and MGD18 deletion strains, a previously generated ISx-Δ*dinP* knockout [59], and an ADP1-ISx strain with a suspected genome rearrangement were constructed as described above. Self-targeting spacer cassettes were created by BsmBI GGA as described above. AZT-resistant colonies were further characterized for mutations in the *tdk-kanR* cassette versus spacer acquisition by patching to determine if they remained Kan resistant and by PCR and Sanger sequencing, if necessary. Oligonucleotides used to conduct these experiments are provided in **Table 3.1**.

RESULTS

Golden Transformation

A. baylyi ADP1 is highly competent for DNA uptake during normal growth. Transformation frequencies of $>10^{-2}$ per cell can be achieved for genomic DNA from a donor ADP1 strain with a genetic marker embedded in its chromosome [2], but transformation is less efficient for shorter DNA fragments. Successful integration of a foreign DNA sequence into the genome, which occurs most readily via RecA-mediated recombination, is generally the limiting step for transformation [174]. Therefore, robust protocols for engineering the genome of ADP1 typically require assembling linear double-stranded DNA cassettes with at least 500 base pairs of flanking genomic homology on each end, and adding even longer flanking homology regions further increases the efficiency of transformation [174,175].

Currently, the most common approach used for ADP1 genome engineering relies on PCR assembly of DNA constructs for transformation. The procedure has variously been called overlap extension PCR (OE-PCR), splicing PCR, fusion PCR, or other names [1,156]. In this method, one first generates a marked replacement DNA cassette containing dual positive and negative selection functionalities (e.g., *tdk* and *kanR* genes) flanked by DNA sequences matching the genomic target site. The PCR product is added to a growing culture of ADP1, allowing DNA uptake and homologous recombination to replace the native sequence between the homology regions. Successful transformants are isolated by plating on agar containing the positive selective agent (e.g., kanamycin). In a second step, an unmarked rescue cassette containing flanking homologies to the insertion site is added to accomplish another recombination event with the chromosome. The rescue cassette can be designed to precisely remove the replacement cassette and create a deletion or to insert new genes into the genome at that site. Successful rescue transformants are isolated by plating on agar containing the negative selection agent (e.g., AZT).

Because it relies on multiple rounds of PCR amplification and must stitch together a longer assembly without giving rise to off-target amplicons, using OE-PCR to create DNA constructs can become cumbersome in large projects. To streamline this process, we tested using Golden Gate Assembly (GGA) to create the replacement and rescue cassettes needed for *A. baylyi* genome engineering more efficiently and quickly. Fundamentally, GGA involves the use of plasmids or PCR fragments encoding parts that are flanked by recognition sites for a type IIS restriction enzyme. The restriction enzyme

cut sites are positioned to generate compatible overhangs between fragments. Multiple DNA pieces can then be efficiently assembled in a designed order in a one-pot reaction that contains both a restriction enzyme and DNA ligase.

For our method, which we dubbed Golden Transformation (GT), we use GGA to combine one or more DNA parts with the genome homology regions required for *A. baylyi* genome engineering (**Fig. 3.1**). In our design, just two PCR reactions that amplify the flanking ~1-kb homology regions from genomic DNA are needed. Each PCR contains a primer that adds terminal BsaI and a BsmBI restriction sites. For the first "replacement" step of genome modification, these flanking regions are combined with the *tdk-KanR* cassette from plasmid pBTK622 in a BsaI GGA reaction. This DNA product is used to transform ADP1 to insert the *tdk-KanR* cassette in place of any genes that are to be deleted. For the second "rescue" step, the same two flanking PCR fragments amplified from the genome are used in a GGA reaction. If a deletion is desired, the two PCR products can be joined together with only a 4-bp scar remaining in a BsmBI GGA reaction. Alternatively, since the BsaI and BsmBI overhangs can be designed to be compatible with GGA-based toolkits of genetic parts [164,176], one can create and insert a transcriptional unit or a larger construct into the genome in place of the *tdk-KanR* cassette.

Testing Golden Transformation

We first tested the efficiency of GT using "replacement" DNA assemblies that included the *tdk-kanR* cassette and flanking genome homology regions (**Fig. 3.2**). OE-

PCR product was used as a benchmark for the expected maximum efficiency that would be possible with the GT procedure. It yielded transformation frequencies of 1.7×10^{-4} per cell, which were similar to previous tests of this construct [59]. Adding GGA buffer containing heat-inactivated T7 ligase and BsaI to the OE-PCR product only reduced transformation by a factor of 1.8-fold [1.3–2.3] (95% confidence interval). Therefore, we concluded that it was not necessary to purify GGA reactions before they were added to cells for transformation.

Golden Transformation using standard 5-hour GGA reactions with three input DNA pieces yielded transformation frequencies that were reduced by roughly two orders of magnitude relative to the OE-PCR benchmark representing the expectation for 100% efficient DNA assembly. Due to the high transformability of *A. bayli* ADP1, this still represents thousands of transformants per standard GT assay. There was very little difference in transformation frequencies whether T4 or T7 ligase was used in GGA. Shortening the steps in the GGA temperature cycling protocol to a shorter 1-hour program reduced the transformation frequency by a factor of 6.9, consistent with successful DNA assembly limiting GT. Interestingly, GT of a larger five-part assembly was roughly as efficient as GT of the three-part assembly. The ability to create this type of complex cassette and integrate it into the *A. baylyi* genome in a single step demonstrates a further advantage of the GT strategy over using OE-PCR for DNA assembly.

We next tested whether one could compensate for the reduced efficiency of GT by performing "puddle" transformations. In this setup, *A. baylyi* and DNA are combined

on a filter that is transferred to an agar surface for cell growth, as opposed to the normal transformation procedure in which DNA is added to cells in a liquid culture. The puddle procedure improved the transformation frequency for each of three GGA reactions that were tested by a similar amount. The group-wise increase by a factor of 5.3 [4.1–6.8] (95% confidence interval) matched the 5-fold higher amount of DNA per cell. The absolute yield of successfully transformed cells is also higher in puddle transformations, despite a reduction in the total number of cells present after growth by a factor of 1.9, on average, compared to the transformation assays in liquid culture.

Tn-Seq to determine ADP1 Gene Essentiality in LB

To demonstrate how Golden Transformation can be used to rapidly construct deletions, we applied it to an *A. baylyi* ADP1 genome streamlining project. The ADP1 genome sequence is annotated with 3305 protein-coding genes and 100 RNA genes [23]. Previously, a single-gene deletion strain collection was constructed by replacing individual open reading frames with the *tdk-kanR* cassette [177]. It identified 499 proteins as essential and 2593 as dispensable for viability in a minimal succinate growth medium. We used transposon sequencing (Tn-Seq) to understand which ADP1 genes are required for robust growth in rich medium (LB), so that we could identify regions of the genome that could be deleted while preserving growth under a broad range of conditions.

Sequencing >4 million reads in the Tn-Seq library, corresponding to 59,757 unique insertion sites, enabled us to confidently classify 2,871 proteins as either essential or dispensable to ADP1 fitness in LB (**Fig. 3.3A, Table S1**). Comparing this set to the

knockout collection, 238 proteins were classified as essential in both experiments, 261 were essential for viability in minimal medium only, 108 were essential for fitness in rich medium only, and 2231 were dispensable in both experiments. Assigning ADP1 proteins to Clusters of Orthologous Groups (COGs) showed that certain functions were overrepresented within each of these categories (**Fig. 3.3B**). As expected, proteins with translation, ribosomal structure and biogenesis functions (J) are most common in the shared essential group, and proteins in the shared dispensable group commonly have an unknown function (S) or are not even assigned to a COG (X). Amino acid transport and metabolism functions (E) predominate among those proteins that are essential for growth in minimal medium and dispensable for maintaining fitness in rich medium, presumably because many biosynthesis pathways are unnecessary when amino acids are supplied as nutrients. Genes essential for fitness in LB yet dispensable in minimal succinate medium are largely involved in energy production and conversion (C). This observation is consistent with ADP1 needing to utilize a more diverse pool of compounds for energy generation in rich medium.

Multiple-gene-deletion Strains

We used information about gene essentiality from the knockout collection and our Tn-seq results to design 55 multi-gene deletions covering a total of 2.13 Mb (59.4%) of the 3.59 Mb ADP1-ISx genome (**Fig. 3.4A**, **Table 3.2**). In addition to spanning as many nonessential genes as possible, some of the regions targeted for deletion included one or a few genes that were predicted to be essential in one or both experiments. We expected

that some of these genes might be conditionally dispensable (i.e., able to be deleted if other nearby genes involved in the same process were removed at the same time). For example, the DNA methylase in a restriction-modification system is essential for cell viability unless its corresponding restriction enzyme is also deleted. Similar interactions can arise in metabolic pathways that have toxic intermediates and in other coupled cellular processes.

We attempted construction of the 55 planned ADP1-ISx derivatives with multiple-gene-deletions using Golden Transformation in rich medium (LB). Single-gene deletion studies commonly encounter problems with false-positives, such as integration of the selectable marker into one copy of a region of the genome that has been transiently amplified in the target cell [178–180]. To detect these artifacts, we validated that the *tdk-kanR* selection cassette inserted into the genome at the expected site and replaced all copies of the targeted genome region in candidate deletion strains by performing several PCR reactions and, in some cases, by re-sequencing whole genomes (**Fig. 3.7**, **Table 3.2**). In total, 19 multiple-gene-deletion (MGD) strains were successfully created. Collectively, they dispense with 24.6% of the ADP1-ISx genome.

For 10 of the 36 regions that we could not delete—which we designate as "retained genome regions" (RGRs)—we did not obtain any transformants following GT, and there were generally 10- to 100-fold fewer transformants in GTs for the other 26 RGRs compared to those that yielded successful deletions (**Table 3.2**). Of these 26 putative deletions that subsequently failed PCR verification, most appeared to have either one-sided integrations of the *tdk-kanR* cassette with rearrangements or incomplete deletions

on the opposite side or to have large chromosomal amplifications which enabled one copy of the targeted genes to be replaced by the *tdk-kanR* cassette while leaving one or more other copies intact. The *A. baylyi* genome is known to experience gene amplifications at high rates [181,182], which led to false-positives during construction of the single-gene knockout collection [183]. To check for these or other types of false-positives in the MGD strains, we validated a subset via whole-genome sequencing, including as controls two strains that failed the PCR assays. All 14 of the sequenced MGD strains were found to have the expected deletions. Genome sequencing corroborated that the other two strains, which had been transformed with DNA targeting either RGR16 or RGR28 for deletion, had one-sided integrations of the *tdk-kanR* cassette. In each case there was a deletion of only a few hundred bases on the opposite flank of the cassette.

We further examined whether the presence of one or more proteins found to be essential in the minimal-medium single-gene knockout collection or in the rich-medium Tn-Seq experiment was predictive of whether a deletion was ultimately successful (**Fig. 3.4B**). Deleted genome regions were roughly half as likely to contain at least one gene flagged as essential in the single-gene knockout collection as regions that we were unable to remove ($p = 0.014$, respectively, one-tailed Fisher's exact test). There was also a lower chance of having a gene designated essential by the Tn-Seq experiment within a successfully deleted region, but not significantly so ($p = 0.186$). The worse correlation with the Tn-Seq experiment most likely reflects that essentiality as measured by this assay is for retaining high fitness rather than absolute viability for growth.

Growth characteristics of MGD strains

We next tested the viability and growth rates of the deletion strains. They were constructed in LB, so they are all viable in this rich medium. However, nearly all deletion strains had moderate 10-25% reductions in maximum specific growth rate in LB compared to the ancestral ADP1-ISx strain, and the MGD9, MGD14 and MGD15 strains exhibited larger reductions of 50% or more (**Fig. 3.4C**). The magnitudes of these growth defects were not predictable from the overall characteristics of the deletion. For example, there was not a significantly different growth rate depending on whether a deletion contained a protein classified as essential in the Tn-Seq experiment or in the deletion collection ($p = 0.29$ and $p = 0.40$, respectively, two-tailed *t*-tests). There was also no consistent correction between the size of the genome region deleted and the growth rate of a strain ($p = 0.31$, for a non-zero slope when bootstrap resampling by strain). Finally, only 2 of the 19 deletion strains (MGD9 and MGD15) showed no growth when cultured in minimal succinate medium for 72 hours.

Reprogramming the native CRISPR-Cas system of *A. baylyi* ADP1

The ADP1 genome encodes a complete type I-Fa CRISPR-Cas system [184]. We tested whether this system was functional by reprogramming it to target a foreign DNA sequence (**Fig. 3.4A**). First, we replaced the entire native CRISPR array of spacers that is proximal to the Cas-operon with the *tdk-kanR* cassette (**Fig. 3.5A**). We refer to an *A. bayli* strain with this modification as "CRISPR-ready" because one or more custom spacers can be re-inserted into the native location to reprogram the CRISPR-ready strain with a transformation that uses AZT counterselection against *tdk* (**Fig. 3.5B**).

The protospacer adjacent motif (PAM) for type I CRISPR-Cas systems has been reported to be a downstream GG dinucleotide [185], including for the subtype I-F system of *Acinetobacter baumanii* [186]. This is equivalent to a CC located 5' of the spacer sequence on the strand that it matches in the target site DNA sequence. We found that matches for some of the native *A. bayli* spacers to *Acinetobacter*-related phages supported this PAM sequence using the CRISPRTarget software [187]. Therefore, we tested the ability of an *A. bayli* ADP1-ISx strain reprogrammed with a single spacer matching a site in a spectinomycin-resistance gene sequence (*specR*) with this PAM to prevent transformation of this genetic marker from genomic DNA isolated from a strain with the *specR* gene integrated into its genome (**Fig. 3.5C**). As controls, we tested the wild-type and CRISPR-ready derivative of this strain and also two variants in which the CRISPR array was re-targeted to *specR* gene sequences without the correct PAM. The properly re-targeted strain reduced transformation frequency by a factor of 10^5, whereas the transformation frequency was unchanged in the other strains, indicating that the native *A. baylyi* CRISPR-Cas system is functional and highly efficient.

Assuring and securing deletions with a CRISPR-Lock

We next considered how CRISPR spacers targeting the chromosomal regions removed in the multiple-gene deletion strain could be used for two purposes to further the ADP1 genome streamlining project. First, we tested whether it was possible to add a deletion-targeting spacer to the genome for assurance: to verify whether a gene has been successfully eliminated from a candidate MGD strain. Second, if one has two MGD

strains, one with a spacer targeting its own deletion and the other that still retains the *tdk-kanR* cassette replacing its deletion, then genomic DNA from the latter can be used to transform the former to readily combine the deletions in one strain. In this case, the deletion-targeting spacer serves as a "CRISPR-Lock" that prevents restoration of the genes originally deleted in the first strain from the genomic DNA of the second strain, which is expected to be strongly selected for during this transformation if it causes a large growth defect.

We tested the assurance scheme by attempting to reprogram CRISPR-ready strain derivatives with self-targeting spacers (T2 and T18) aimed at sequences that were removed in two of the mutants from the deletion collection, MGD2 and MGD18, respectively, or with the first spacer from the native CRISPR array (N1) as an off-target control (**Fig. 3.6**). We would expect to not be able to transform a strain that does not have a deletion with a self-targeting spacer because this will lead to cleavage of its own genome and cell death. However, escape mutants can evolve that have spontaneous mutations inactivating the *tdk* gene that give them resistance to AZT without replacement of the *tdk-kanR* cassette in the step in which we add the spacer to a CRISPR-ready strain. The frequency of ~10^{-6} of background AZT-resistant (AZT^R) escape mutants makes it difficult to determine whether the self-targeted region was present in the genome based on measuring differences in transformation frequencies of the different spacers alone. Therefore, we screened 10 colonies for whether they maintained kanamycin resistance (Kan^R) as a quick proxy for whether putative spacer transformants were actually escape mutants.

CRISPR-Ready variants of wild-type ADP1-ISx, MGD2, and MGD18 were constructed as detailed above (**Fig. 3.5**), after first removing the *tdk-kanR* cassette from the deleted regions in the latter two strains using GT with the flanking PCR homology segments (**Fig. 3.1**). Of 10 AZT^R colonies isolated after transformation of each of these strains with the control N1 spacer construct, all 10 were all kanamycin sensitive (Kan^S) in all three cases, indicating 100% successful integration of this off-target spacer (**Fig. 3.6**). When we transformed wild-type ADP1-ISx with either of the self-targeting spacers (T2 or T18) all 10 AZT^R colonies characterized were now Kan^R, which the expected result because these spacers should be lethal in the absence of their cognate deletions. In the case of the T18 construct, we found that all 10 colonies had a 551-bp deletion inactivating the *tdk* gene. Most AZT^R colonies isolated after transformation of MGD2 and MGD18 with spacers targeting the regions deleted in each strain were Kan^S (8/10 and 9/10, respectively). Sequencing confirmed that these isolates had proper spacer integration. The one exception for MGD18 contained a 1-bp deletion in the *tdk* gene. It is also possible for AZT^R colonies to be Kan^S if they incorporate a mutated copy of the spacer that no longer matches the genome, but we did not observe that outcome. These results show that there significant difference in the chances of getting AZT^R colonies that are sensitive or resistant to kanamycin ($p = 2.6 \times 10^{-8}$, Fisher's exact test of combined T2 and T18 results) that is enough to make this assay diagnostic for assuring that a gene has been successfully deleted.

Successful MGD2+T2 and MGD18+T18 transformants from this experiment were used to test the CRISPR-Lock approach. We transformed these strains with

genomic DNA from the MGD17 strain still containing the *tdk-kanR* replacement cassette. The MGD17 deletion is separated by less than 900 bp from the one in MGD18 (**Table 3.2**). In each case, 3/3 *kanR* colonies screened had successfully added the new deletion and retained the original one, as determined by PCR and Sanger sequencing. However, we also were able to create the same double deletions by transforming MGD2 and MGD18 strains that did not have the self-targeting T2 or T18 spacers with the same efficiency of 3/3 *kanR* colonies screened having both deletions. Though it was not necessary to block re-incorporation of the deleted segments with the CRISPR-Lock when making these double deletion strains, we expect that it will be increasingly important to prevent re-acquisition of deleted regions as more and more deletions are combined and result in strains with progressively worse fitness as this genome streamlining process continues.

DISCUSSION

We developed the Golden Transformation procedure to address limitations in the current methods used to assemble DNA cassettes for modifying the genome of the naturally competent bacterium *A. baylyi* ADP1. The ease of Golden Gate assembly combined with the very high transformability of unpurified DNA from these reactions into *A. baylyi* enables a highly streamlined procedure that requires fewer steps and fewer oligonucleotides than PCR assembly methods. We illustrated the utility of Golden Transformation by generating a collection of 19 derivatives of the transposon-free *A.*

baylyi ADP1-ISx strain that each have one additional very large chromosomal deletion spanning 20,920 to 183,258 base pairs and 19 to 172 genes.

Determining how removing large regions of a genome impacts viability and growth is a first step toward achieving a streamlined genome. The success of 19 of our 55 attempts to erase large regions from the *A. baylyi* genome under rich medium conditions depended to some extent on whether they contained genes judged to be essential for viability or maintaining high fitness when they were removed or inactivated one-at-a-time. There was a higher correlation between deletion success for genes absolutely required for viability in a single-gene knockout collection in minimal medium than there was for genes that were needed to maintain high fitness in a TnSeq library constructed in rich medium. These single-gene results were not able to completely predict whether a larger deletion containing multiple genes would be viable, and the exceptions can give interesting information about the overall systems-organization of the ADP1 genome. For example, there can be synthetic lethals in which deletion of two or more genes leads to a loss in viability even when none of the genes is essential on their own [188,189]. Multiple-gene deletions may also synergistically perturb gene expression and metabolic pathways in ways that create deleterious imbalances that are bigger than for single-gene deletions [79,91].

Only 2 of the 19 *A. baylyi* large deletions lost their ability to grow in minimal media. This result was expected for one of these strains (MGD15) because it should be incapable of pyrimidine biosynthesis due to deletion of the *pyrF* gene. Supplementation with uracil was able to restore growth of this deletion strain in minimal succinate

medium, showing that *pyrF* could be used as an auxotrophic selection marker in ADP1 as it is in many other bacterial species [190]. The other such deletion (MGD9) did not contain any genes that led to inviability in minimal medium when they were deleted one-at-a-time in the knockout strain collection. ADP1 encodes two isocitrate dehydrogenase isozymes encoded by the *ACIAD1187* and *ACIAD1190* genes [23]. Both of these genes were simultaneously deleted in this strain, so they likely result in synthetic lethality in the minimal medium that has succinate as the sole carbon source.

Conversely, 4 of the 19 MGD strains exhibited growth in minimal medium despite the removal of genes that were essential in the single-gene knockout collection. MGD8 included the deletion of four putative essential genes (*lysS*, *cysD*, *cysN* and ACIAD1056). MGD13 showed growth despite lacking two genes found to be essential in the single-gene deletion collection (*terD* and *ACIAD1965*). MGD7 deleted the ATP-binding subunit of an iron transporter (*ACIAD0969*) and a protein of unknown function (*ACIAD1000*). Finally, MGD19 included a putative gene of unknown function (*ACIAD3600*) essential for growth in minimal media.

Many successful deletions—in 10 of the 19 MGD strains—included genes found to be essential for fitness in the Tn-Seq experiment that were not essential for viability in the deletion collection in minimal medium. These genes included several sugar transferases (*ACIAD0084- ACIAD0091*), a lipid metabolism gene (*fadB*), an outer membrane protein (*ACIAD0697*), an elongation factor (*typA*), a putative hemolysin (*ACIAD0944*), isocitrate dehydrogenase (*idh*), a putative transthyretin-like protein (*ACIAD1188*), a succinylglutamate desuccinylase (*astE*), a transcriptional regulator

(*dcaS*), an efflux pump(*smvA*), uridylyltransferase (*glnD*), and cysteine synthase A (*cysK*), another putative transcriptional regulator (ACIAD1082) and a penicillin-binding protein (*pbpA*). Loss of these genes may contribute to the reduced growth rates of the multiple-gene deletion strains, making them top candidates for adding back one or a few of the removed genes in each segment to maintain high fitness.

The overall worse growth of the MGD strains may be the result from metabolic imbalances that result from genome reduction [86] or deletion of quasi-essential genes [54]. The most growth-detrimental deletions were MGD15 and MGD9. A possible explanation is that among the genes deleted in MGD15 is a gene essential in succinate minimal media (*pyrF*) in addition to malate synthase G (*glcB*), a gene involved in the TCA cycle which is essential for growth in acetate and 2,3-butanediol as sole carbon sources [183]. Similarly, two genes deleted in MGD9, *ACIAD2740* and *ACIAD2741*, were previously determined to be critical for growth in acetate and quinate, respectively [183] and their loss may be particularly disruptive to metabolism.

Several of the other engineered MGDs in this study have some interesting features and may reveal fitness-balancing effects. For instance, MGD1 may emulate the fitness-enhancing interruptions of *per* and *pgi* genes that have been observed in an adaptive laboratory evolution experiment with ADP1 [19]. These beneficial deletions may buffer detrimental effects generated by the rest of the content in the deletion. In this case, the inclusion of *galU* in MGD1 could be counteracting the fitness-enhancing effects of *per* and *pgi* deletions, since a *galU* knockout shows defective growth in acetate, 2,3-butanediol, and quinate [183]. MGD1 may have benefited from the removal of an entire

fimbrial operon which is likely costly to express (*ACIAD0119- ACIAD0123*). MGD5 may have two potent fitness-enhancing deletions: of the type IV fimbrial biogenesis protein FimT) and of the small regulatory RNA AbsR28 [191]. Mutations in the latter gene been observed previously during ADP1 adaptive evolution [19].

Other deletions may illustrate the advantages and disadvantages of concentrating on genome minimization in one environment while not examining robustness to other environments. MGD5 includes several possible detrimental deletions: a stress gene *ACIAD0704* (*mscL* homolog with known functions in hypoosmotic shock and in mechanical, nutritional and oxidative stress), DNA repair gene *mutM* (formamidopyrimidine-DNA glycosylase) and *tonB* (heme/iron acquisition). It is likely that the genome deletion is less robust in terms of both its genome stability and its ability to survive nutrient and physical stresses. MGD11, one of the larger deletions (183 kb and 167 genes), tolerates the somewhat surprising elimination of one copy of the 16S and 23S ribosomal genes, along with Ala and Ile tRNA copies. While it does not delete any essential genes deemed essential for viability in minimal media, it does remove two fitness-essential genes in LB, *cysK* and *dcaS*. In this case, it also seems likely that there could be some fitness balancing effect of removing multiple pathways involved in the degradation of aromatic compounds that are contained within this region (including *pca-qui-pob-hca* genes) [13,15].

Overall, our data show that the impact of single-gene deletions on growth cannot fully predict the results we observe with the simultaneous deletion of multiple genes and that complex fitness-modulating interactions may have important roles reduced-genome

bacteria [183]. Methods that use genome-scale models may better inform the deletion process [192], including how one can relocate of one or a few genes that are needed for viability/fitness from a stretch of chromosome before it is deleted. The *pyrH* and *icd* genes are candidates for this approach. They could be added back to the deleted region during the Golden Transformation rescue step, for example. Perhaps a bigger issue for the genome reduction enterprise is the potential for losing robustness in other environments when genes are deleted from a strain. The extent of this loss of robustness remains to be tested for the MGD strains in future work.

RNA-guided nucleases have revolutionized the field of genome editing among many other applications. Much of this work has focused on CRISPR-Cas9, but there are a large diversity of CRISPR types in bacterial genomes that are less well-characterized [193]. The type I-Fa CRISPR-Cas system of *A. baylyi* is analogous to the type I-Fb CRISPR/Cas system of *A. baumanii*, except the *cas1* gene appears last in the Cas operon [184,186]. *A. baylyi* ADP1 contains three CRISPR arrays: one with 91 spacers adjacent to the Cas operon and two with 21 and 6 spacers located elsewhere in the genome. By using Golden Transformation to first replace the 91-spacer CRISPR array with the *tdk-kanR* dual selection cassette to create a "CRISPR-Ready" strain and then inserting a single synthetic spacer in place of this cassette, we showed that *A. baylyi*'s CRISPR system is active and readily reprogrammable.

Novel spacers may be designed and transformed into a CRISPR-Ready *A. baylyi* strain, opening up this system for various applications. For example, we create a "self-targeting" CRISPR-Lock that tests whether a certain gene has been successfully removed

from the chromosome and prevents its re-acquisition. Deletions in the strains we created could be combined in future work by adding genomic DNA isolated from one deletion strain to a locked version of another deletion strain. This strategy could be particularly effective if self-targeting spacers are added in place of each deleted region so that this protection will also accumulate during genome streamlining.

Although orthogonal dCas9 gene repression has been applied successfully in *A. bayli* ADP1 [194], there was a metabolic cost for expressing this foreign system. The ability to reprogram the endogenous CRISPR/Cas type I-F system opens up the possibility of repurposing it for programmable transcriptional repression (analogous to dCas9) via *cas3* knockout [195]. These applications of the native CRISPR system broaden the opportunities for genome-wide screens and combinatorial pathway engineering in this strain.

In summary, we demonstrated new methods for more facile and assured genome editing in *Acinetobacter baylyi*. We used Golden Transformation to create 19 large deletions that outline a future roadmap for significantly reducing its genome complexity to make it a simpler and more predictable metabolic and genome engineering platform. This method also simplifies assembling multiple DNA parts in a one-pot reaction before they are added to the genome, opening up new possibilities for constructing and testing combinatorial libraries. Additionally, by showing that the native *A. baylyi* CRISPR-Cas system is active we also open up further possibilities for restricting transformation of DNA that might reverse our engineered deletion and for re-purposing this system for genetic control. These developments leverage and further improve upon *A. baylyi*'s remarkable utility as a chassis organism that is highly naturally transformable.

FUNDING

This work was supported by the National Institutes of Health [R00-GM087550]; the Defense Advanced Research Projects Agency [HR0011-15-C0095]; the National Science Foundation BEACON Center for the Study of Evolution in Action [DBI-0939454]; and the National Science Foundation [CBET-1554179]; and the Welch Foundation [F-1979-20190330].

ACKNOWLEDGEMENTS

We thank Ellen Neidle for sharing the puddle transformation protocol.

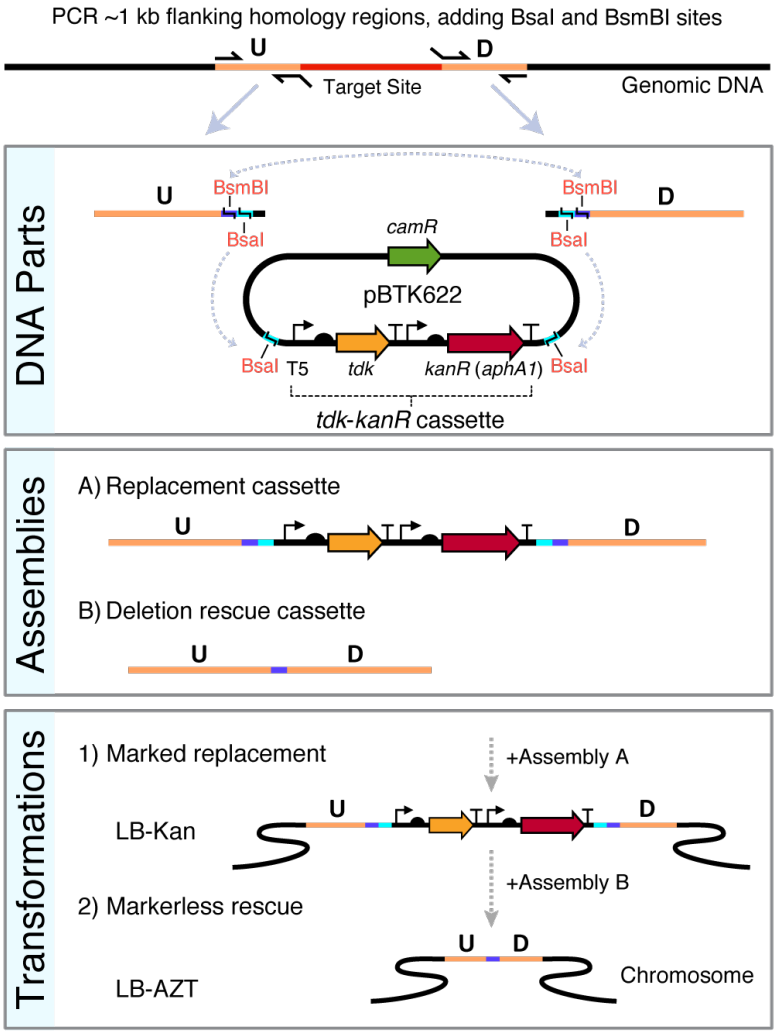

Figure 3.1. Golden Transformation method for ADP1 genome engineering.

Two PCR reactions are performed for upstream (U) and downstream (D) genomic target flanks that add terminal BsaI and BsmBI type-IIs restriction sites as depicted. The two PCR products can then be combined via BsaI Golden Gate assembly (GGA) with the selection cassette to form a replacement DNA or combined with one another and optionally with additional genetic parts (not shown) via BsmBI GGA to form a rescue cassette. The positive-negative selection cassette (*tdk-kanR*) is maintained on the high-copy pBTK622 plasmid that has an origin that does not replicate in *A. baylyi*. The first GGA reaction is added to an *A. baylyi* culture and then plated on LB-Kan to select for transformants with the replacement cassette integrated into the genome. Then, transformation of the second assembly reaction with counterselection on LB-AZT is used to move the unmarked deletions/additions encoded on the rescue cassette into the genome.

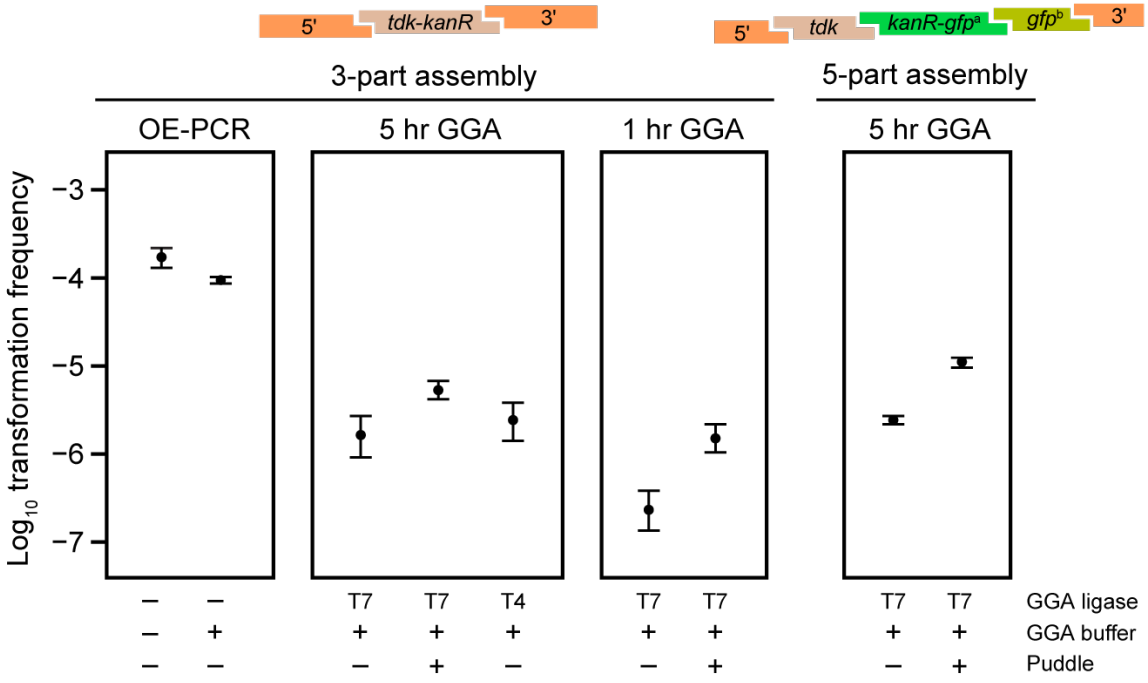

Figure 3.2. Golden Transformation can achieve high genome editing rates.

Transformation of two different cassettes constructed in 3-part or 5-part Golden Gate assembly (GGA) reactions were compared by counting colonies obtained on LB-Kan versus LB agar. Transformations of a purified DNA sample constructed by overlap-extension PCR (OE-PCR) at a concentration that corresponds to 100% efficient assembly of the 3-part GGA reaction and this OE-PCR sample with GGA buffer added to it were included for comparison. The effects of "short" (1 hr) or "long" (5 hr) GGA thermocycling programs, using T4 versus T7 ligase in GGA reactions, and performing "puddle" transformations that concentrate cells and DNA by combining them on a filter placed on an agar surface rather than in a liquid culture were also tested.

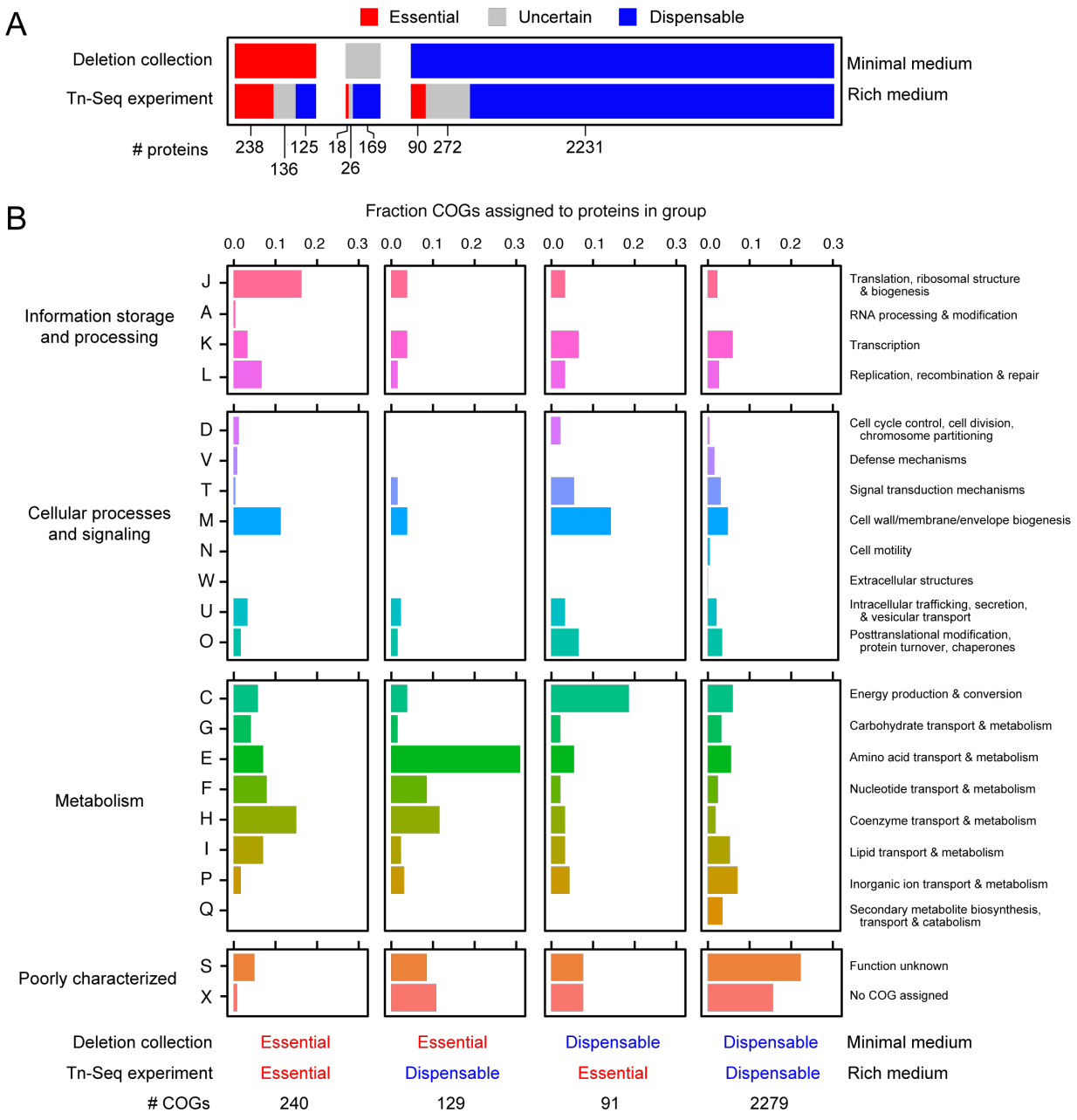
A
Essential
Uncertain
Dispensable
Deletion collection
Tn-Seq experiment
Minimal medium
Rich medium
proteins
238
136
125
18
26
169
90
272
2231
B
Fraction COGs assigned to proteins in group
0.0
0.1
0.2
0.3
Information storage and processing
J
A
K
L
Translation, ribosomal structure & biogenesis
RNA processing & modification
Transcription
Replication, recombination & repair
Cellular processes and signaling
D
V
T
M
N
W
U
O
Cell cycle control, cell division, chromosome partitioning
Defense mechanisms
Signal transduction mechanisms
Cell wall/membrane/envelope biogenesis
Cell motility
Extracellular structures
Intracellular trafficking, secretion, & vesicular transport
Posttranslational modification, protein turnover, chaperones
Metabolism
C
G
E
F
H
I
P
Q
Energy production & conversion
Carbohydrate transport & metabolism
Amino acid transport & metabolism
Nucleotide transport & metabolism
Coenzyme transport & metabolism
Lipid transport & metabolism
Inorganic ion transport & metabolism
Secondary metabolite biosynthesis, transport & catabolism
Poorly characterized
S
X
Function unknown
No COG assigned
Deletion collection
Essential
Essential
Dispensable
Dispensable
Minimal medium
Tn-Seq experiment
Essential
Dispensable
Essential
Dispensable
Rich medium
COGs
240
129
91
2279

Figure 3.3.

Figure 3.3. *A. bayli* ADP1 protein essentiality

(A) Comparison of the essentiality of protein-coding genes between two experiments. The deletion collection determined which strains lacking a single protein remained viable in a defined minimal succinate medium (minimal medium) [183]. The Tn-Seq experiment from the current study examined the representation of transposon insertions in different genes within populations of mutants cultured in LB (rich medium), which reveals proteins that are essential for fitness in this environment. Colored bars correspond to the numbers of proteins that were judged as essential, dispensable, or uncertain (meaning either untested or an ambiguous). The horizontal bars are broken up to show the overlap of proteins classified into each category between experiments. (B) Functions of proteins classified as various combinations of essential or dispensable across the two experiments. Each column shows the breakdown of Clusters of Orthologous Groups (COG) functions predicted for genes that were definitively classified in each experiment (not in the uncertain category).

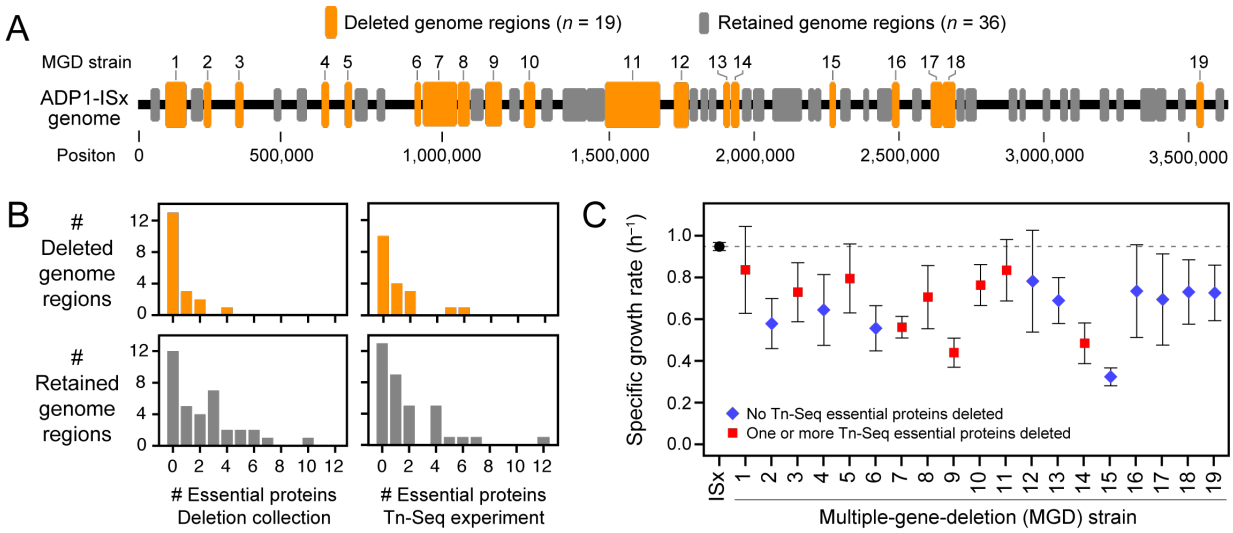

Figure 3.4. Dispensability of *A. baylyi* genome regions targeted for deletion and growth rates of multiple-gene deletion strains.

(A) Regions targeted for deletion from the ADP1 chromosome. Successful deletions that resulted in multiple-gene deletion (MGD) strains are numbered and displayed in orange. Deletions that could not be constructed are shown in gray. (B) Gene essentiality in deleted and retained regions. Histograms show the breakdown of the 19 deleted regions and the 36 retained regions by how many proteins they included that were classified as essential in the deletion collection or the Tn-Seq experiment. (C) Maximum specific growth rates of MGD strains in LB determined from fitting growth curves monitored by optical density in a microplate reader. Shapes and colors of symbols indicate whether any essential proteins from the Tn-Seq experiment were included in that deletion. Error bars are 95% confidence intervals.

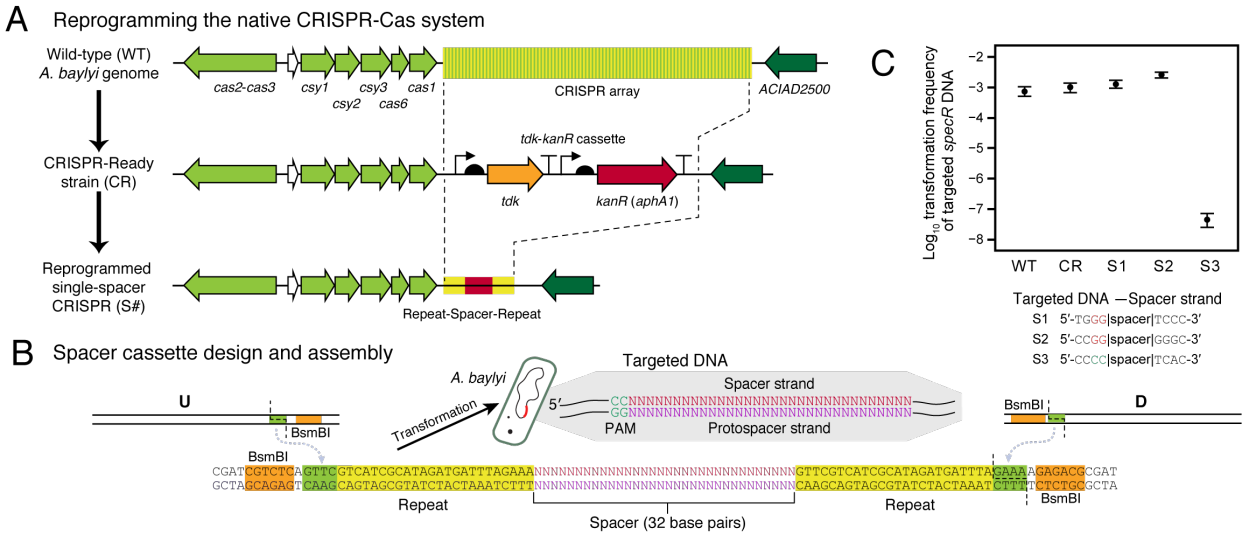

Figure 3.5. The native *A. baylyi* ADP1 CRISPR-Cas system is active and can be retargeted.

(A) Scheme for reprogramming the *A. baylyi* CRISPR array. First, a "CRISPR-Ready" strain is created by performing a Golden Transformation that replaces the entire native spacer array with a *tdk-kanR* cassette. Then, a second Golden Transformation can be used to add a rescue cassette that contains one or more designed spacers under control of the native gene expression signals. (B) Single-spacer replacement cassette design. Synthetic double-stranded DNA encoding the spacer and surrounding repeats is combined with PCR products corresponding to the flanking genome homology upstream (U) and downstream (D) using BsmBI Golden Gate assembly. The inset shows the DNA sequence that is targeted for cleavage with the protospacer adjacent motif (PAM) typical of type-I CRISPR-Cas systems. (C) Reprogrammed CRISPR-Cas system restricts transformation of foreign DNA. Frequencies of transformants of genomic DNA from *A. baylyi* ADP1 donor that has an integrated spectinomycin resistance gene (*specR*) were used to judge whether targeting a spacer to this sequence in a recipient strain prevented its acquisition. WT is wild type ADP1-ISx and CR is the CRISPR-Ready derivative of this strain. S1 and S2 are controls with spacers that match the *specR* sequence but in incorrect PAM contexts. S3 an on-target *specR* spacer with the correct PAM. Error bars are estimated 95% confidence intervals.

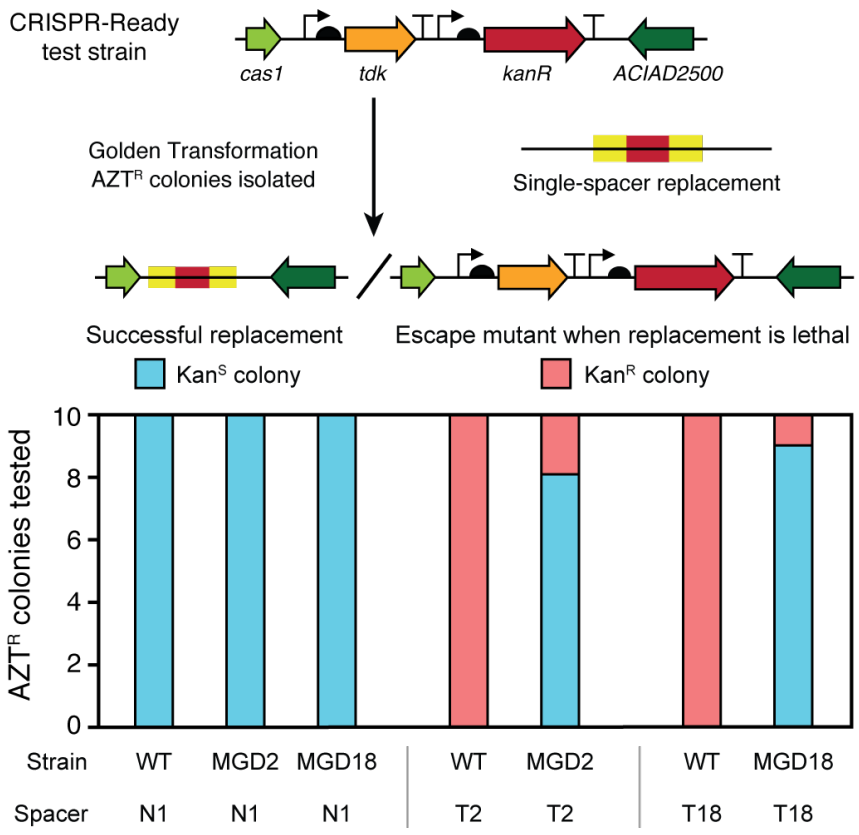

Figure 3.6. Self-targeting spacers can be used to assure deletions and create a CRISPR-Lock.

CRISPR-Ready variants of wild-type ADP1-ISx (WT) and two multiple gene deletions strains (MGD2 and MGD18) were transformed with different spacers to assess the presence of a sequence located within the putatively deleted region. N1 added back the first spacer sequence from the native CRISPR array. It serves as a control because it does not target any sequence in the *A. baylyi* genome. T2 and T18 are spacers that match sequences in the ADP1-ISx genome that are within the regions deleted in the corresponding MGD strains. For each strain-spacer combination tested, 10 AZT^R colonies were isolated after transformation with the single-spacer replacement DNA. Successful integration of the spacer can only occur if the targeted region is not present in the recipient strain's genome. It results in these AZT^R colonies also becoming Kan^S. If integration of the spacer is lethal, then AZT^R colonies are expected to have mutations that inactivate the *tdk* gene and remain Kan^R, as illustrated in the upper panel. Strains with successful spacer integrations from the MGD2+T2 and MGD18+T18 transformations have a "CRISPR-Lock" in their genomes that can prevent re-acquisition of the deleted regions when combining multiple deletions in further stages of the genome streamlining project.

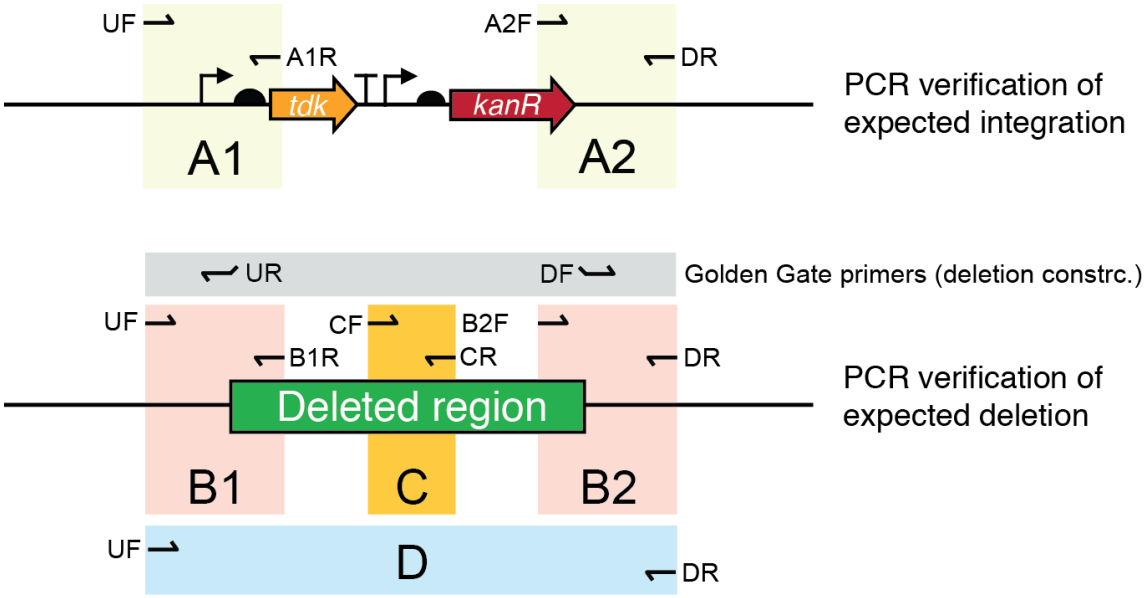

Figure 3.7. Deletion mutant validation PCR reactions.

PCR reactions A, B, C and D, were employed during deletion mutant validation. Successful amplification from A1 or A2 PCRs served as positive indication of integration of marker cassette at the correct loci, whereas amplification from B1, B2 or C reactions gave indication of failed deletion and possible genome rearrangement.

Experiment	Description	Oligo	Sequence
multiple-gene deletion	*tdk-kanR* primer for all PCR A1	A1R	ACATCAGAGATTTTGAGACACAACGTG
multiple-gene deletion	*tdk-kanR* primer for all PCR A2	A2F	GATTAAATTCCAACATGGATGC
multiple-gene deletion	RGR1	UF	TCACACCGACCTTACCG
multiple-gene deletion	RGR1	UR	ATGCGGTCTCACGTTCGTCTCAGACCGTACAACCCATCCAATCG
multiple-gene deletion	RGR1	DF	GCATGGTCTCAGCTGCGTCTCAGGTCGTTTGTCAATTTTCAGGATCG
multiple-gene deletion	RGR1	DR	CTTGCGGCATAAATAAAGC
multiple-gene deletion	DGR1 (MGD1)	UF	GATGGAAAACATATTTCAGATAAGC
multiple-gene deletion	DGR1 (MGD1)	UR	ATGCGGTCTCACGTTCGTCTCAGACCCGAATCTATATAAGAGGGATTATTTTT
multiple-gene deletion	DGR1 (MGD1)	DF	GCATGGTCTCAGCTGCGTCTCAGGTCATCATTGCTATATCGAACGAAC
multiple-gene deletion	DGR1 (MGD1)	DR	CAGGATCAACTTCACCTGC
multiple-gene deletion	DGR1 (MGD1)	B1R	CTCTTCAGGGCTGATCTG
multiple-gene deletion	DGR1 (MGD1)	B2F	GGTGCAAAGGTACGAGC
multiple-gene deletion	RGR2	UF	CATTTACTGAAAGCAACAAGC
multiple-gene deletion	RGR2	UR	ATGCGGTCTCACGTTCGTCTCAGACCGCTCAATCAATTTAGCAAAGC
multiple-gene deletion	RGR2	DF	GCATGGTCTCAGCTGCGTCTCAGGTCTCGTCCTCTGTACTTCCG
multiple-gene deletion	RGR2	DR	GCTTCAAACCCATACCG
multiple-gene deletion	DGR2 (MGD2)	UF	GGCTGCTTTCAAGAATATAAGC
multiple-gene deletion	DGR2 (MGD2)	UR	ATGCGGTCTCACGTTCGTCTCAGACCGCCTTATGATGCTGAAGC
multiple-gene deletion	DGR2 (MGD2)	DF	GCATGGTCTCAGCTGCGTCTCAGGTCCTTCTGCCGTTGGTCG
multiple-gene deletion	DGR2 (MGD2)	DR	TTTCATCAAGTACATAAACACGC
multiple-gene deletion	DGR2 (MGD2)	B1R	ATCGCCATTGGTTGC
multiple-gene deletion	DGR2 (MGD2)	B2F	GAGGACGAATTGATGAAGC
multiple-gene deletion	DGR3 (MGD3)	UF	CTTGGTGATTACCGAGCC
multiple-gene deletion	DGR3 (MGD3)	UR	ATGCGGTCTCACGTTCGTCTCAGACCCTTTCTTTCAACTCAGCATAAGC
multiple-gene deletion	DGR3 (MGD3)	DF	GCATGGTCTCAGCTGCGTCTCAGGTCTGCTTATGAGTTACATGATTGG
multiple-gene deletion	DGR3 (MGD3)	DR	GTGGTGTAATAAAATGCTTGC
multiple-gene deletion	DGR3 (MGD3)	B1R	CACAGATCCAATACCAGC
multiple-gene deletion	DGR3 (MGD3)	B2F	GAGCAGTTCTCAAGAATGC
multiple-gene deletion	RGR3	UF	CTGACCAAACTGACCACG
multiple-gene deletion	RGR3	UR	ATGCGGTCTCACGTTCGTCTCAGACCAGTTAAAGCAGTATCGCCG
multiple-gene deletion	RGR3	DF	GCATGGTCTCAGCTGCGTCTCAGGTCCTAATTTCGATAAAAATCCACTGC
multiple-gene deletion	RGR3	DR	CGATAAAATGATGCTGGC
multiple-gene deletion	RGR3	B1R	CGTTTAATTGTGCATTTGC
multiple-gene deletion	RGR3	B2F	CCGATCTTAAAAGGAGATGC
multiple-gene deletion	RGR3	CF	AGCGTTGAATCCTGTGC

Table 3.1. Oligonucleotides used in this study

Experiment	Description	Oligo	Sequence
multiple-gene deletion	RGR3	CR	CTTGCTGCGTTTGAAGC
multiple-gene deletion	RGR4	UF	GACCACATTCGAGATGC
multiple-gene deletion	RGR4	UR	ATGCGGTCTCACGTTCGTCTCAGACCCGCGTATAAATAATTTTGCC
multiple-gene deletion	RGR4	DF	GCATGGTCTCAGCTGCGTCTCAGGTCCCAGGGTCTAACAGTGAGC
multiple-gene deletion	RGR4	DR	TGTCATCAAAATCAATTCTGC
multiple-gene deletion	RGR4	B1R	ATCAACCATCCTAATCATGC
multiple-gene deletion	RGR4	B2F	GTCCCGGTGAGTATTTACG
multiple-gene deletion	RGR4	CF	GCGAATTTGACCATCG
multiple-gene deletion	RGR4	CR	GCAGGCAGTAGGTCAGC
multiple-gene deletion	DGR4 (MGD4)	UF	TCAATCTCAATTCATGTCGC
multiple-gene deletion	DGR4 (MGD4)	UR	ATGCGGTCTCACGTTCGTCTCAGACCTTCACAAAGTTGATTAAATCCG
multiple-gene deletion	DGR4 (MGD4)	DF	GCATGGTCTCAGCTGCGTCTCAGGTCGGTTTACATTTTTACCGTTGC
multiple-gene deletion	DGR4 (MGD4)	DR	CTTTGTCCATGGTAGGTGC
multiple-gene deletion	DGR4 (MGD4)	B1R	ACTTCGGTGGTAACTCTGC
multiple-gene deletion	DGR4 (MGD4)	B2F	CAATTGGATCAAGTTTAGTTGC
multiple-gene deletion	DGR5 (MGD5)	UF	CCTGTACAGAAGAAACATTTGC
multiple-gene deletion	DGR5 (MGD5)	UR	ATGCGGTCTCACGTTCGTCTCAGACCCTGCATTGTTTATTCTATCACCAG
multiple-gene deletion	DGR5 (MGD5)	DF	GCATGGTCTCAGCTGCGTCTCAGGTCTTCGAGCTCAATACATCATATC
multiple-gene deletion	DGR5 (MGD5)	DR	ATCCTGAATGTCATCATATTCG
multiple-gene deletion	DGR5 (MGD5)	B1R	AACCATGTGTGTTTGCG
multiple-gene deletion	DGR5 (MGD5)	B2F	GTTGTCGCAACATCTGC
multiple-gene deletion	RGR5	UF	ACGTTGTCGAGCATACG
multiple-gene deletion	RGR5	UR	ATGCGGTCTCACGTTCGTCTCAGACCGGCTCTTTGTTATGCTATGC
multiple-gene deletion	RGR5	DF	GCATGGTCTCAGCTGCGTCTCAGGTCGTAGAGATTTCCAAATTGTGC
multiple-gene deletion	RGR5	DR	GCGCTGAAAACAAAGTACG
multiple-gene deletion	RGR5	B1R	AACTGATACATCTCGGTTGC
multiple-gene deletion	RGR5	B2F	TTCAGCATTAGAAAACTGTGC
multiple-gene deletion	RGR5	CF	AGCAAGTTCATCATTACGC
multiple-gene deletion	RGR5	CR	TCCACATAGCATCACATGA
multiple-gene deletion	RGR6	UF	GATATACTGGTTCCTGATTGGTC
multiple-gene deletion	RGR6	UR	ATGCGGTCTCACGTTCGTCTCAGACCCAGAAAAGACTCAAAATAAGATGTAATTC
multiple-gene deletion	RGR6	DF	GCATGGTCTCAGCTGCGTCTCAGGTCGGGCAATACTATTAGTGATGAAGC
multiple-gene deletion	RGR6	DR	CGTCTGCACCAACTACAGC
multiple-gene deletion	RGR6	B1R	CGCCAGTCATTGATGC
multiple-gene deletion	RGR6	B2F	CGTATAGAGCAAGGCCG

Table 3.1. Oligonucleotides used in this study (continued)

Experiment	Description	Oligo	Sequence
multiple-gene deletion	DGR6 (MGD6)	UF	GCCATACTGGAACCGC
multiple-gene deletion	DGR6 (MGD6)	UR	ATGCGGTCTCACGTTCGTCTCAGACCCCAAAATCAACTTTTTCACC
multiple-gene deletion	DGR6 (MGD6)	DF	GCATGGTCTCAGCTGCGTCTCAGGTCTTTATCATCGCTTCCAAAATG
multiple-gene deletion	DGR6 (MGD6)	DR	GTTTCTGGCTGGTCTGC
multiple-gene deletion	DGR6 (MGD6)	B1R	TCTTGATTGCCTCTGCC
multiple-gene deletion	DGR6 (MGD6)	B2F	GTGTAGTTGGAGAGAAAGTTGC
multiple-gene deletion	DGR7 (MGD7)	UF	AAACTAATAGTCAGAGCGTGAGTTATAC
multiple-gene deletion	DGR7 (MGD7)	UR	TGATTTGAATTGGAGGCTGGGTATTTTTTACCTCTCTACCACCAATA
multiple-gene deletion	DGR7 (MGD7)	UR2 (rescue)	AAGCGTGACCAGACCACTTATTTTTTACCTCTCTACCACCAATA
multiple-gene deletion	DGR7 (MGD7)	DF	TTCTAAGCATGCGGAGCTGGAGTGGTCTGGTCACGCTT
multiple-gene deletion	DGR7 (MGD7)	DF2 (rescue)	TATTTTTTACCTCTCTACCACCAATAAGTGGTCTGGTCACGCTT
multiple-gene deletion	DGR7 (MGD7)	DR	CTCAATGAACGCTTGGAG
multiple-gene deletion	DGR8 (MGD8)	UF	ATGCCATAGATGTTTACCTGC
multiple-gene deletion	DGR8 (MGD8)	UR	ATGCGGTCTCACGTTCGTCTCAGACCGAGCACCACAACAACTGC
multiple-gene deletion	DGR8 (MGD8)	DF	GCATGGTCTCAGCTGCGTCTCAGGTCGATTATAAATCACGTTTTGGC
multiple-gene deletion	DGR8 (MGD8)	DR	GACGCTTTACAATTTCTTCG
multiple-gene deletion	DGR8 (MGD8)	B1R	TCCAATTGCATGACAGC
multiple-gene deletion	DGR8 (MGD8)	B2F	AATGGATAAAGGACTTCGC
multiple-gene deletion	DGR8 (MGD8)	CF1	CACCAGTCAGATCTTATTAGCC
multiple-gene deletion	DGR8 (MGD8)	CR1	GATAAAGATAATTATAGGCGAACC
multiple-gene deletion	DGR8 (MGD8)	CF2	GGACTGGAAAGGAGATGC
multiple-gene deletion	DGR8 (MGD8)	CR2	TGCTTGCTCAAGATTGC
multiple-gene deletion	RGR7	UF	GATTGATCGAAATGCCG
multiple-gene deletion	RGR7	UR	ATGCGGTCTCACGTTCGTCTCAGACCCGCTTTGAATATGGTGC
multiple-gene deletion	RGR7	DF	GCATGGTCTCAGCTGCGTCTCAGGTCCATCGGATTAAATGGATCG
multiple-gene deletion	RGR7	DR	GCATGTCAGATCAGATCG
multiple-gene deletion	RGR7	B1R	CCAATGTTGTTCTTAATACCG
multiple-gene deletion	RGR7	B2F	GCGGTATCTTATTGATCCG
multiple-gene deletion	RGR7	CF	TTATTTGGCAAATAGACATGG
multiple-gene deletion	RGR7	CR	ATGTCACAAACAATCACGCT
multiple-gene deletion	DGR9 (MGD9)	UF	GATAGGTGCATTACATGATGC
multiple-gene deletion	DGR9 (MGD9)	UR	ATGCGGTCTCACGTTCGTCTCAGACCCTCATTGGGAAAAACGTG
multiple-gene deletion	DGR9 (MGD9)	DF	GCATGGTCTCAGCTGCGTCTCAGGTCCTAAGTTCGGCTCAGATTC
multiple-gene deletion	DGR9 (MGD9)	DR	GCAAGATTGACTCCACG
multiple-gene deletion	DGR9 (MGD9)	B1R	AGTCCAACATGACCGC

Table 3.1. Oligonucleotides used in this study (continued)

Experiment	Description	Oligo	Sequence
multiple-gene deletion	DGR9 (MGD9)	B2F	TGGACAATGCTGATACTGC
multiple-gene deletion	RGR8	UF	ATATTGCGTTGCCAGC
multiple-gene deletion	RGR8	UR	ATGCGGTCTCACGTTCGTCTCAGACCGATCCAAATCATACCAATCG
multiple-gene deletion	RGR8	DF	GCATGGTCTCAGCTGCGTCTCAGGTCGGTTGATTCAGCAGGC
multiple-gene deletion	RGR8	DR	CATAAGCCAAAGCATTTGC
multiple-gene deletion	RGR8	B1R	ATCAGCTTGGTGATATTGC
multiple-gene deletion	RGR8	B2F	GCACGCTATTGTTACGC
multiple-gene deletion	RGR8	CF	GCGGTAAATGCAGAATGTATC
multiple-gene deletion	RGR8	CR	ATGAACATCTTCAACATTTAGATCAAG
multiple-gene deletion	DGR10 (MGD10)	UF	CGTGGCAATTATCAATGC
multiple-gene deletion	DGR10 (MGD10)	UR	ATGCGGTCTCACGTTCGTCTCAGACCAACGTTCCACTATTGTTTGTC
multiple-gene deletion	DGR10 (MGD10)	DF	GCATGGTCTCAGCTGCGTCTCAGGTCGATCAGAATTAGCAAAAGCAAC
multiple-gene deletion	DGR10 (MGD10)	DR	GTTCTGGAAATGGAAGGC
multiple-gene deletion	DGR10 (MGD10)	B1R	GGATTCTCTGTCAAAGTTCG
multiple-gene deletion	DGR10 (MGD10)	B2F	AGACAACGCGAGTCTGC
multiple-gene deletion	RGR9	UF	CCTTACCCATATCTTTCATGC
multiple-gene deletion	RGR9	UR	ATGCGGTCTCACGTTCGTCTCAGACCTTGATGTCGTAGATCAGAACG
multiple-gene deletion	RGR9	DF	GCATGGTCTCAGCTGCGTCTCAGGTCCGTTCTTATGCAGATGCG
multiple-gene deletion	RGR9	DR	GCCACCATAACTGATTGC
multiple-gene deletion	RGR10	UF	AATCATGAAAACTCCATCTGG
multiple-gene deletion	RGR10	UR	ATGCGGTCTCACGTTCGTCTCAGACCGGTGGTTATCATTTAGCACG
multiple-gene deletion	RGR10	DF	GCATGGTCTCAGCTGCGTCTCAGGTCCTTGGCAAGGTTGTGC
multiple-gene deletion	RGR10	DR	AGAGATGGTTTCAAGTGCG
multiple-gene deletion	RGR11	UF	GGTATAAGGATCCCAAATGC
multiple-gene deletion	RGR11	UR	ATGCGGTCTCACGTTCGTCTCAGACCAGAGATGGTTTCAAGTGCG
multiple-gene deletion	RGR11	DF	GCATGGTCTCAGCTGCGTCTCAGGTCGTAGCTTTAAAAGGCAATATGC
multiple-gene deletion	RGR11	DR	GAGTATGTGCTTAGCTGGTGC
multiple-gene deletion	RGR11	B1R	CAAATCACCATGACCTGC
multiple-gene deletion	RGR11	B2F	GCTACGAATGACCAGACG
multiple-gene deletion	RGR11	CF	AACCACATTGGGACTGC
multiple-gene deletion	RGR11	CR	AATGGGTTTATCAATTTGGC
multiple-gene deletion	DGR11 (MGD11)	UF	AAATTACCTTGATAGCCTGCA
multiple-gene deletion	DGR11 (MGD11)	UR	TGATTTGAATTGGAGGCTGGGAAGCTCTAAAAGGCAGCAT
multiple-gene deletion	DGR11 (MGD11)	UR2 (rescue)	GTAAATTATTTTATAGAAATTTTAAGAGAAATATAAGCTCTAAAAGGCAGCAT
multiple-gene deletion	DGR11 (MGD11)	DF	TTCTAAGCATGCGGAGCTGGATATTTCTCTTAAAATTTCTATAAAATAATTTAC

Table 3.1. Oligonucleotides used in this study (continued)

Experiment	Description	Oligo	Sequence
multiple-gene deletion	DGR11 (MGD11)	DF2 (rescue)	AAGCTCTAAAAGGCAGCATATATTTCTCTTAAAATTTCTATAAAATAATTTAC
multiple-gene deletion	DGR11 (MGD11)	DR	CGATCTTATGCAGGAGGTAC
multiple-gene deletion	DGR12 (MGD12)	UF	GCTTTGGACTTGCTCTCG
multiple-gene deletion	DGR12 (MGD12)	UR	ATGCGGTCTCACGTTCGTCTCAGACCCAATTGACGTATTGACGC
multiple-gene deletion	DGR12 (MGD12)	DF	GCATGGTCTCAGCTGCGTCTCAGGTCTCGATGATCGTTGATATAGAGG
multiple-gene deletion	DGR12 (MGD12)	DR	TTCGGATGAATTTTAAACGC
multiple-gene deletion	DGR12 (MGD12)	B1R	GCACTTCCCATGATTGC
multiple-gene deletion	DGR12 (MGD12)	B2F	CGTTTATCTTCAGAATGTCCG
multiple-gene deletion	RGR12	UF	GCCCTGAAACATTAACATCG
multiple-gene deletion	RGR12	UR	ATGCGGTCTCACGTTCGTCTCAGACCGGGACGCTTTATCATGTGC
multiple-gene deletion	RGR12	DF	GCATGGTCTCAGCTGCGTCTCAGGTCGCAGGCATACGGTATTCAAC
multiple-gene deletion	RGR12	DR	GCAAATACTACGGCAAATACTACG
multiple-gene deletion	RGR12	B1R	CCAATGCTGAAATTATGGC
multiple-gene deletion	RGR12	B2F	CACACTGATATGTTTACCTTCG
multiple-gene deletion	RGR13	UF	GTCATGATCTACCCCTTGC
multiple-gene deletion	RGR13	UR	ATGCGGTCTCACGTTCGTCTCAGACCTTCTTGAACTCGAATTTGAACTG
multiple-gene deletion	RGR13	DF	GCATGGTCTCAGCTGCGTCTCAGGTCGCGGAATAACAGATGAAGAA
multiple-gene deletion	RGR13	DR	CTTTAGCGTGGATAAGCG
multiple-gene deletion	RGR13	B1R	GCGTGATTATGTTACATAGGC
multiple-gene deletion	RGR13	B2F	TTGACCTTGAGCAAGTAAGC
multiple-gene deletion	RGR14	UF	GCAGTTGGGAAAGAATCG
multiple-gene deletion	RGR14	UR	ATGCGGTCTCACGTTCGTCTCAGACCTGGTGAGAAATCTGATGGC
multiple-gene deletion	RGR14	DF	GCATGGTCTCAGCTGCGTCTCAGGTCTTACAGGCGTTGAGAATAGC
multiple-gene deletion	RGR14	DR	AAGAACAACAGGCAGGC
multiple-gene deletion	RGR14	B1R	AACTTAATGCTGCATCTTGC
multiple-gene deletion	RGR14	B2F	AACGTTATCCCATAGGTGC
multiple-gene deletion	RGR14	CF	CAATCGTGACCTGTACTGC
multiple-gene deletion	RGR14	CR	GCATGGAAGACAGCGC
multiple-gene deletion	DGR13 (MGD13)	UF	ATGTTTGCATATAAGGGGC
multiple-gene deletion	DGR13 (MGD13)	UR	ATGCGGTCTCACGTTCGTCTCAGACCCGCGTACTCAATCAATGC
multiple-gene deletion	DGR13 (MGD13)	DF	GCATGGTCTCAGCTGCGTCTCAGGTCTGCTTGGTCGATGTCG
multiple-gene deletion	DGR13 (MGD13)	DR	CTTCACGAACCAAAGTCG
multiple-gene deletion	DGR13 (MGD13)	B1R	GTTTGGGAAACATTTGTCG
multiple-gene deletion	DGR13 (MGD13)	B2F	GCTTCTATAGGTTCACCATGC
multiple-gene deletion	DGR13 (MGD13)	CF	ATCGAAAGTGTTGTGATCG

Table 3.1. Oligonucleotides used in this study (continued)

Experiment	Description	Oligo	Sequence
multiple-gene deletion	DGR13 (MGD13)	CR	AAGCTTGCGATGTGTGC
multiple-gene deletion	DGR14 (MGD14)	UF	TGGTGTACTTTCATCATCAGC
multiple-gene deletion	DGR14 (MGD14)	UR	ATGCGGTCTCACGTTCGTCTCAGACCGTTCTTCGGTAGGAGAGGC
multiple-gene deletion	DGR14 (MGD14)	DF	GCATGGTCTCAGCTGCGTCTCAGGTCGAATTTGTTTAGAGCTGGGC
multiple-gene deletion	DGR14 (MGD14)	DR	CTTAATGATGAACACAGCATCG
multiple-gene deletion	DGR14 (MGD14)	B1R	GCTGATGTGTGCTGTAAGC
multiple-gene deletion	DGR14 (MGD14)	B2F	CTAGCGCAGAAGACTTGC
multiple-gene deletion	RGR15	UF	GACTCAAGTACATTCATTCCG
multiple-gene deletion	RGR15	UR	ATGCGGTCTCACGTTCGTCTCAGACCCAAACCAGTCGCAAGC
multiple-gene deletion	RGR15	DF	GCATGGTCTCAGCTGCGTCTCAGGTCTTTCATGTTGAATAGAATATCTTGC
multiple-gene deletion	RGR15	DR	ACCACGACTACGATTTGC
multiple-gene deletion	RGR15	B1R	GGTGTTTGCTGATTCTGC
multiple-gene deletion	RGR15	B2F	ATGGGTAAAGCCTAGAGTTACG
multiple-gene deletion	RGR16	UF	GTTGGAGAGTTTGAATCGC
multiple-gene deletion	RGR16	UR	ATGCGGTCTCACGTTCGTCTCAGACCCATATTGTCCAAGAAGACTTTGC
multiple-gene deletion	RGR16	DF	GCATGGTCTCAGCTGCGTCTCAGGTCCAAATTGATTACACGGAAAGC
multiple-gene deletion	RGR16	DR	GACGAATCAAGTCATTAATCG
multiple-gene deletion	RGR16	B1R	TCTAACCAACTGAGCTATAGGC
multiple-gene deletion	RGR16	B2F	ATGCCAATGCTTTCTCG
multiple-gene deletion	RGR17	UF	TTCAAGTAGAACAGCTAAAGGC
multiple-gene deletion	RGR17	UR	ATGCGGTCTCACGTTCGTCTCAGACCTAGCGTATATGGCATGAGG
multiple-gene deletion	RGR17	DF	GCATGGTCTCAGCTGCGTCTCAGGTCAAACTTTCATCTCTTTTTAATCAGG
multiple-gene deletion	RGR17	DR	CTATTGGGCTGTTTAAATAGAGTATG
multiple-gene deletion	RGR18	UF	GATCAAAGAGGCCATGC
multiple-gene deletion	RGR18	UR	ATGCGGTCTCACGTTCGTCTCAGACCACAACCAGATACCAAGTCACC
multiple-gene deletion	RGR18	DF	GCATGGTCTCAGCTGCGTCTCAGGTCTTTCTGAAAACTCTGGACG
multiple-gene deletion	RGR18	DR	GTCACATTACTCTGGGTGC
multiple-gene deletion	RGR18	B1R	TATCTTTGAAGCACGTAGTATGC
multiple-gene deletion	RGR18	B2F	TTGATGAATACAGCCAGAGC
multiple-gene deletion	RGR18	CF	CAGCCTGCTCTTAATACGC
multiple-gene deletion	RGR18	CR	ACCCTCGCTAATGTTTGC
multiple-gene deletion	RGR19	UF	TTGATGAATACAGCCAGAGC
multiple-gene deletion	RGR19	UR	ATGCGGTCTCACGTTCGTCTCAGACCTGCTTCTGAGGCTTAATGC
multiple-gene deletion	RGR19	DF	GCATGGTCTCAGCTGCGTCTCAGGTCAACTGTTGCAGAAATTCAGC
multiple-gene deletion	RGR19	DR	AGAGATGCACATTAGAGATTGC

Table 3.1. Oligonucleotides used in this study (continued)

Experiment	Description	Oligo	Sequence
multiple-gene deletion	DGR15 (MGD15)	UF	CCGAATGCGACTTATGC
multiple-gene deletion	DGR15 (MGD15)	UR	ATGCGGTCTCACGTTCGTCTCAGACCTTCAAATGTTAAATGGATATTCG
multiple-gene deletion	DGR15 (MGD15)	DF	GCATGGTCTCAGCTGCGTCTCAGGTCCGATGACCTCTTTAATTATTGC
multiple-gene deletion	DGR15 (MGD15)	DR	CAACTGATGGTTAATTATGTCG
multiple-gene deletion	DGR15 (MGD15)	B1R	CACGAGACATATAGGTTGTACG
multiple-gene deletion	DGR15 (MGD15)	B2F	ACGTAATGCTTTACCAGGC
multiple-gene deletion	DGR15 (MGD15)	CF	CATCCAACATTGCCTGC
multiple-gene deletion	DGR15 (MGD15)	CR	GTATTATTGTTGCCTTAGACGC
multiple-gene deletion	RGR20	UF	CAACCATTTTGCTTGACG
multiple-gene deletion	RGR20	UR	ATGCGGTCTCACGTTCGTCTCAGACCCTCGATTTTATGTCTACCGC
multiple-gene deletion	RGR20	DF	GCATGGTCTCAGCTGCGTCTCAGGTCTGAACCTGATCAAAACATCG
multiple-gene deletion	RGR20	DR	TTGCAGGTGGTTTTAATGC
multiple-gene deletion	RGR20	B1R	CTGAGCGTTTTGGTGC
multiple-gene deletion	RGR20	B2F	GGTATCCCACTTGCTGC
multiple-gene deletion	RGR20	CF	TCATCATTCAACATAACATATTGC
multiple-gene deletion	RGR20	CR	GAGAGATTGTTCTACATGCTCG
multiple-gene deletion	RGR21	UF	CGAAGTAAATATCGTAATCACTGC
multiple-gene deletion	RGR21	UR	ATGCGGTCTCACGTTCGTCTCAGACCCGCAAGTTGTTTGACACG
multiple-gene deletion	RGR21	DF	GCATGGTCTCAGCTGCGTCTCAGGTCTGGTCAGTTGGCTTATCG
multiple-gene deletion	RGR21	DR	AATATACATGCCTATTAAAGTCGC
multiple-gene deletion	RGR21	B1R	TGTTGAGCGTGATGTACG
multiple-gene deletion	RGR21	B2F	CACTGTAGAACAAATGCTGC
multiple-gene deletion	RGR21	CF	GAAGTTTCTGAACATACTGGC
multiple-gene deletion	RGR21	CR	CATTGACACGTTTGTAGTCG
multiple-gene deletion	RGR22	UF	CGAACCGCTTACAGGC
multiple-gene deletion	RGR22	UR	ATGCGGTCTCACGTTCGTCTCAGACCACAGGCTGTGAGTTATCCG
multiple-gene deletion	RGR22	DF	GCATGGTCTCAGCTGCGTCTCAGGTCTTTAAACACAACCCATTATTCG
multiple-gene deletion	RGR22	DR	AAAGAAGCCGATCTAAATGC
multiple-gene deletion	RGR22	B1R	CGAACGTATTGTGCTCG
multiple-gene deletion	RGR22	B2F	AACAGCACCATATCGGC
multiple-gene deletion	RGR22	CF	CAGGATGACACCTATGACG
multiple-gene deletion	RGR22	CR	TATGGAAAGCGTGTGAGC
multiple-gene deletion	DGR16 (MGD16)	UF	GATGTCAGTGATCATAATGAACG
multiple-gene deletion	DGR16 (MGD16)	UR	ATGCGGTCTCACGTTCGTCTCAGACCCATTTGCAGCAATGGATT
multiple-gene deletion	DGR16 (MGD16)	DF	GCATGGTCTCAGCTGCGTCTCAGGTCGCTATCGTCATGTTTATGCC

Table 3.1. Oligonucleotides used in this study (continued)

Experiment	Description	Oligo	Sequence
multiple-gene deletion	DGR16 (MGD16)	DR	CGACCTTCATTCGTGC
multiple-gene deletion	DGR16 (MGD16)	B1R	TCTTATGGTAGTTATGATGGTGC
multiple-gene deletion	DGR16 (MGD16)	B2F	GCCTAAAGCTATTTGGAAGC
multiple-gene deletion	RGR23	UF	AAACGCTCAGTGAAGAAGC
multiple-gene deletion	RGR23	UR	ATGCGGTCTCACGTTCGTCTCAGACCTCAATGCCAGTGTCAGC
multiple-gene deletion	RGR23	DF	GCATGGTCTCAGCTGCGTCTCAGGTCCATTACCACTGGAGTTATAAACG
multiple-gene deletion	RGR23	DR	GACTGCCCAAGTTTATCG
multiple-gene deletion	DGR17 (MGD17)	UF	GTACATCTGTCACCCCTGC
multiple-gene deletion	DGR17 (MGD17)	UR	ATGCGGTCTCACGTTCGTCTCAGACCCTTCCTCTATCCAATCAGTGC
multiple-gene deletion	DGR17 (MGD17)	DF	GCATGGTCTCAGCTGCGTCTCAGGTCTACAACCATAATCTGAGCAGAG
multiple-gene deletion	DGR17 (MGD17)	DR	CGATTCCAATTTTACTCTGTTC
multiple-gene deletion	DGR17 (MGD17)	B1R	CGCCTTTACCATCACG
multiple-gene deletion	DGR17 (MGD17)	B2F	ACGAAAACAGTCACTTGATTG
multiple-gene deletion	DGR18 (MGD18)	UF	GCCGTGCTCTTATTATATCCAC
multiple-gene deletion	DGR18 (MGD18)	UR	ATGCGGTCTCACGTTCGTCTCAGACCATAAGTAGTGATATTTTGTATACGCAC
multiple-gene deletion	DGR18 (MGD18)	DF	GCATGGTCTCAGCTGCGTCTCAGGTCCGATGAGCAAATGATCAAAG
multiple-gene deletion	DGR18 (MGD18)	DR	GTTTCATATACTTTACGTGATGTCG
multiple-gene deletion	DGR18 (MGD18)	B1R	CATAGTACTGACCTCGGAAGC
multiple-gene deletion	DGR18 (MGD18)	B2F	TACCCTGGTCATGGACG
multiple-gene deletion	RGR24	UF	TCAGAAATAATTCAATAAAACAGGG
multiple-gene deletion	RGR24	UR	ATGCGGTCTCACGTTCGTCTCAGACCGTTTTATCGACAGTGTGGGC
multiple-gene deletion	RGR24	DF	GCATGGTCTCAGCTGCGTCTCAGGTCGAATCTGCTCTATTTGTGAAGC
multiple-gene deletion	RGR24	DR	GTGGGCAATAGTCAATGG
multiple-gene deletion	RGR24	B1R	CGTTACTAGTCAAATGGATGC
multiple-gene deletion	RGR24	B2F	AATTCATCTAGACGCTGTGC
multiple-gene deletion	RGR25	UF	CTTCAATTCACTAATGGATTTGC
multiple-gene deletion	RGR25	UR	ATGCGGTCTCACGTTCGTCTCAGACCTAAAGTCTTATCAAAGAATAGGGC
multiple-gene deletion	RGR25	DF	GCATGGTCTCAGCTGCGTCTCAGGTCTCTATACTCCAAATTTTTCTATCAATGC
multiple-gene deletion	RGR25	DR	GTACACTCAATGCAGTCGC
multiple-gene deletion	RGR25	B1R	GTGGGCAATAGTCAATGG
multiple-gene deletion	RGR25	B2F	AGCTCTTACTGCAATGTCG
multiple-gene deletion	RGR26	UF	CAGTTCAAGACGGACGC
multiple-gene deletion	RGR26	UR	ATGCGGTCTCACGTTCGTCTCAGACCGGATGAAGGGAAATAAAATGAC
multiple-gene deletion	RGR26	DF	GCATGGTCTCAGCTGCGTCTCAGGTCGCGTCATTTATTATGCGG
multiple-gene deletion	RGR26	DR	GCAGGTGAAATTGCCG

Table 3.1. Oligonucleotides used in this study (continued)

Experiment	Description	Oligo	Sequence
multiple-gene deletion	RGR26	B1R	AATGTTACTGATTTGGAGCG
multiple-gene deletion	RGR26	B2F	CGTTACTCACCACCACG
multiple-gene deletion	RGR26	CF	TGGAAGAATTAACTTTTTCGC
multiple-gene deletion	RGR26	CR	CCATCTATTTCGAGGCC
multiple-gene deletion	RGR27	UF	GCAATTTGAATACCTAACGG
multiple-gene deletion	RGR27	UR	ATGCGGTCTCACGTTCGTCTCAGACCTCGAATAATAACTAAAGAACCGC
multiple-gene deletion	RGR27	DF	GCATGGTCTCAGCTGCGTCTCAGGTCAATATCATCACATTGGAATGC
multiple-gene deletion	RGR27	DR	ATATAAGGACGGGTAGGCG
multiple-gene deletion	RGR28	UF	GGATTAAGTGTCGGTTTGAGC
multiple-gene deletion	RGR28	UR	ATGCGGTCTCACGTTCGTCTCAGACCGTAATCCAGCAACATTGGC
multiple-gene deletion	RGR28	DF	GCATGGTCTCAGCTGCGTCTCAGGTCAGCTGACTTGGAAACGC
multiple-gene deletion	RGR28	DR	GCTGTATGGTTGACAACACG
multiple-gene deletion	RGR28	B1R	TCGGTGACACAGGACG
multiple-gene deletion	RGR28	B2F	CAACCTGAATAGCCTTGC
multiple-gene deletion	RGR29	UF	ATAATCCATCACCTCGGC
multiple-gene deletion	RGR29	UR	ATGCGGTCTCACGTTCGTCTCAGACCCAGCAGAGTGTTGTTTAGAGC
multiple-gene deletion	RGR29	DF	GCATGGTCTCAGCTGCGTCTCAGGTCCCTTTAGGCGGATGATATGC
multiple-gene deletion	RGR29	DR	GCGAGTATAAATGTGTCCTGC
multiple-gene deletion	RGR29	B1R	CTAAGTTCGCTATCGAGGC
multiple-gene deletion	RGR29	B2F	CTCATATCTATGGATAAATTGGC
multiple-gene deletion	RGR29	CF	CAGTTGATTGGATATACAGGC
multiple-gene deletion	RGR29	CR	ACATGACAATGAACATTACTGC
multiple-gene deletion	RGR30	UF	TGCAGAAATTGAGATTACCG
multiple-gene deletion	RGR30	UR	ATGCGGTCTCACGTTCGTCTCAGACCGCATTGGTTAATCTAATGGC
multiple-gene deletion	RGR30	DF	GCATGGTCTCAGCTGCGTCTCAGGTCGCAGAAACACAAATAGCGC
multiple-gene deletion	RGR30	DR	GCTGCCAATGTGATGC
multiple-gene deletion	RGR30	B1R	CAAGCAAGATCATTTGGC
multiple-gene deletion	RGR30	B2F	GTTGTTGGTAAGATCACTTTAGC
multiple-gene deletion	RGR31	UF	AGCGGACACACAGAAGC
multiple-gene deletion	RGR31	UR	ATGCGGTCTCACGTTCGTCTCAGACCCATTTCTAGAGTTACTTGAAGTATAGC
multiple-gene deletion	RGR31	DF	GCATGGTCTCAGCTGCGTCTCAGGTCGTATCCATCTTTAGACAACATGC
multiple-gene deletion	RGR31	DR	CTGCATAAGTTTTGTAGTAGGTAGC
multiple-gene deletion	RGR32	UF	CAACACCCAAGAAACAAAGTC
multiple-gene deletion	RGR32	UR	ATGCGGTCTCACGTTCGTCTCAGACCGTGACTATCATCCATCATATAGGAG
multiple-gene deletion	RGR32	DF	GCATGGTCTCAGCTGCGTCTCAGGTCCTGGCTCAATAGGTAATGGC

Table 3.1. Oligonucleotides used in this study (continued)

Experiment	Description	Oligo	Sequence
multiple-gene deletion	RGR32	DR	CAGATGTGGAAAGCTATGC
multiple-gene deletion	RGR32	B1R	CTGGATGGAATCACCG
multiple-gene deletion	RGR32	B2F	TGGTGTCTGGATCAATGC
multiple-gene deletion	RGR32	CF	TTCAACACTAATATAATCGCTCG
multiple-gene deletion	RGR32	CR	GAAATTGGTAACTGTGACCG
multiple-gene deletion	RGR33	UF	GCTGTGCTACCTGTAAACG
multiple-gene deletion	RGR33	UR	ATGCGGTCTCACGTTCGTCTCAGACCGTTACCTTTACCTGCACCG
multiple-gene deletion	RGR33	DF	GCATGGTCTCAGCTGCGTCTCAGGTCGTAAATATCTAAATTGGTGGGC
multiple-gene deletion	RGR33	DR	GTATGGCCTCAGTCAAGC
multiple-gene deletion	RGR33	B1R	AGGACGAGTGCTATGTGC
multiple-gene deletion	RGR33	B2F	CCACCAATGACACCAGC
multiple-gene deletion	RGR33	CF	TCCCTGTCACCACTTGC
multiple-gene deletion	RGR33	CR	GATGTTAGGAGGTTTGTCTGC
multiple-gene deletion	RGR34	UF	GCAACTGTGACAATTTCAGC
multiple-gene deletion	RGR34	UR	ATGCGGTCTCACGTTCGTCTCAGACCTCGTGGTATTGAAGAAGTCG
multiple-gene deletion	RGR34	DF	GCATGGTCTCAGCTGCGTCTCAGGTCGTCACCCGAATCTCTTGC
multiple-gene deletion	RGR34	DR	AATAACGGTTAGAGATTTCACG
multiple-gene deletion	RGR34	B1R	GCCTTTAATGTGATGGAAGC
multiple-gene deletion	RGR34	B2F	AGGTGCAATTAGACTTGGC
multiple-gene deletion	RGR34	CF	GATGGCGTTAAGGATATGC
multiple-gene deletion	RGR34	CR	CCAGGTTCTGATTGGC
multiple-gene deletion	RGR35	UF	CATAGGTCAGTGTTGGACG
multiple-gene deletion	RGR35	UR	ATGCGGTCTCACGTTCGTCTCAGACCCCGATGCTTTATGTTGC
multiple-gene deletion	RGR35	DF	GCATGGTCTCAGCTGCGTCTCAGGTCCCTTCAATTTAATGCAGATACG
multiple-gene deletion	RGR35	DR	CGTCCTGGAATGTATACCG
multiple-gene deletion	RGR35	B1R	GTACGTGGCAATCTGGC
multiple-gene deletion	RGR35	B2F	AATAATAACGTCGGACTGAGC
multiple-gene deletion	RGR35	CF	ACTCAATGCCTGTAACTTGC
multiple-gene deletion	RGR35	CR	CGTGAGTGATGGTGAGC
multiple-gene deletion	DGR19 (MGD19)	UF	CCTCTACCAACACTATAAAGCG
multiple-gene deletion	DGR19 (MGD19)	UR	ATGCGGTCTCACGTTCGTCTCAGACCAATAGCTTTGGGATTGGC
multiple-gene deletion	DGR19 (MGD19)	DF	GCATGGTCTCAGCTGCGTCTCAGGTCCGTTACTCATCCCAATGC
multiple-gene deletion	DGR19 (MGD19)	DR	CGCAGTATGTATGACCGC
multiple-gene deletion	DGR19 (MGD19)	B1R	TTCAACATGAGAACTCCAGC
multiple-gene deletion	DGR19 (MGD19)	B2F	AACATCTCGATCTTTCAACG

Table 3.1. Oligonucleotides used in this study (continued)

Experiment	Description	Oligo	Sequence
multiple-gene deletion	DGR19 (MGD19)	CF	GATTGCTAATGTCATCAGTGC
multiple-gene deletion	DGR19 (MGD19)	CR	GTACGTTGGATTCGTTAATCG
multiple-gene deletion	RGR36	UF	CCCAAAGTGATCTGGC
multiple-gene deletion	RGR36	UR	ATGCGGTCTCACGTTCGTCTCAGACCGTCAATACCGATAATTTTTGC
multiple-gene deletion	RGR36	DF	GCATGGTCTCAGCTGCGTCTCAGGTCTAAATTTATTCTTATTAGGGACGC
multiple-gene deletion	RGR36	DR	CCCCTTCTGCTAATGTCG
multiple-gene deletion	RGR36	B1R	GTTTCACCTAAGTAGTCTTCTGC
multiple-gene deletion	RGR36	B2F	TTTAGTAGCCATGCGAGC
CRISPR ready	CRISPR-Ready strain construction	UF	TTATCTGGAAAAAGTTCGTGTC
CRISPR ready	CRISPR-Ready strain construction	UR	ATGCGGTCTCACGTTCGTCTCAGAACTAGAGTTAAGTCAAAACAAAACCCT
CRISPR ready	CRISPR-Ready strain construction	DF	GCATGGTCTCAGCTGCGTCTCAGAAATATTAAAAGCAACATCGATAAGAT
CRISPR ready	CRISPR-Ready strain construction	DR	ACGCGCATCATACTCACTTA
CRISPR test	S1 spacer cassette	TS	CGATCGTCTCAGTTCGTCATCGCATAGATGATTTAGAAACTGGCGATGAGCGAAA TGTAGTGCTTACGTTGGTTCGTCATCGCATAGATGATTTAGAAAAGAGACGCGAT
CRISPR test	S1 spacer cassette	BS	ATCGCGTCTCTTTTCTAAATCATCTATGCGATGACGAACCAACGTAAGCACTACA TTTCGCTCATCGCCAGTTTCTAAATCATCTATGCGATGACGAACTGAGACGATCG
CRISPR test	S2 spacer cassette	TS	CGATCGTCTCAGTTCGTCATCGCATAGATGATTTAGAAACAAAATCGCGCCGAAG GATGTCGCTGCCGACTGTTCGTCATCGCATAGATGATTTAGAAAAGAGACGCGAT
CRISPR test	S2 spacer cassette	BS	ATCGCGTCTCTTTTCTAAATCATCTATGCGATGACGAACAGTCGGCAGCGACATC CTTCGGCGCGATTTTGTTTCTAAATCATCTATGCGATGACGAACTGAGACGATCG
CRISPR test	S3 spacer cassette	TS	CGATCGTCTCAGTTCGTCATCGCATAGATGATTTAGAAATGGAGAGAGCGAGATT CTCCGCGCTGTAGAAGGTTCGTCATCGCATAGATGATTTAGAAAAGAGACGCGAT
CRISPR test	S3 spacer cassette	BS	ATCGCGTCTCTTTTCTAAATCATCTATGCGATGACGAACCTTCTACAGCGCGGAG AATCTCGCTCTCTCCATTTCTAAATCATCTATGCGATGACGAACTGAGACGATCG
CRISPR-lock	N1 spacer cassette	TS	CGATCGTCTCAGTTCGTCATCGCATAGATGATTTAGAAAGACACCAAAGGTAATA AAGCTATGAAAGAATAGTTCGTCATCGCATAGATGATTTAGAAAAGAGACGCGAT
CRISPR-lock	N1 spacer cassette	BS	ATCGCGTCTCTTTTCTAAATCATCTATGCGATGACGAACTATTCTTTCATAGCTT TATTACCTTTGGTGTCTTTCTAAATCATCTATGCGATGACGAACTGAGACGATCG
CRISPR-lock	T2 spacer cassette	TS	CGATCGTCTCAGTTCGTCATCGCATAGATGATTTAGAAATGCAGATGCAGTAAGA CTTGTTTGCTCCCAGAGTTCGTCATCGCATAGATGATTTAGAAAAGAGACGCGAT
CRISPR-lock	T2 spacer cassette	BS	ATCGCGTCTCTTTTCTAAATCATCTATGCGATGACGAACTCTGGGAGCAAACAAG TCTTACTGCATCTGCATTTCTAAATCATCTATGCGATGACGAACTGAGACGATCG
CRISPR-lock	T18 spacer cassette	TS	CGATCGTCTCAGTTCGTCATCGCATAGATGATTTAGAAAATTAAACCGCCTGGAG GCTGATCTGTAATCCTGTTCGTCATCGCATAGATGATTTAGAAAAGAGACGCGAT
CRISPR-lock	T18 spacer cassette	BS	ATCGCGTCTCTTTTCTAAATCATCTATGCGATGACGAACAGGATTACAGATCAGC CTCCAGGCGGTTTAATTTTCTAAATCATCTATGCGATGACGAACTGAGACGATCG

Table 3.1. Oligonucleotides used in this study

Oligo abbreviations: U, upstream; D, downstream; F, forward; R, reverse; TS, top strand; BS, bottom strand. Deletion verification primers are for PCR reactions labeled A1, A2, B1, B2, C as shown in Figure 3.7. Bases highlighted in red serve to mark the spacer region within the spacer cassettes used for CRISPR tests.

Strain	Region	Deletion	MS growth	Start	End	CDS	GT colonies	GT puddle	PCR verifications	WGS
NA	RGR1	failed	NA	31264	60558	39	0	0	NA	NA
MGD1	DGR1	successful	Yes	78742	149394	61	> 100	NA	AB	verified
NA	RGR2	failed	NA	164334	203760	40	0	0	NA	NA
MGD2	DGR2	successful	Yes	208059	233310	19	> 50	NA	AB	verified
MGD3	DGR3	successful	Yes	311062	338405	26	> 50	NA	AB	verified
NA	RGR3	failed	NA	439602	466208	25	7	NA	(A)(B)(C)	NA
NA	RGR4	failed	NA	516394	551625	28	< 5	NA	(A)(B)(C)	NA
MGD4	DGR4	successful	Yes	599230	624488	23	> 100	NA	AB	verified
MGD5	DGR5	successful	Yes	676281	701116	23	< 5	NA	AB	NA
NA	RGR5	failed	NA	710503	752357	38	0	1	(B)(C)	NA
NA	RGR6	failed	NA	783695	810856	29	< 5	NA	(A)(B)	NA
MGD6	DGR6	successful	Yes	907554	934015	16	> 100	NA	(A)B	verified
MGD7	DGR7	successful	Yes	934985	1050978	107	NA	NA	CD	verified
MGD8	DGR8	successful	Yes	1051557	1093313	43	< 5	NA	(A)(B)C	NA
NA	RGR7	failed	NA	1095916	1140655	39	< 5	1	(A)(B)(C)	NA
MGD9	DGR9	successful	No	1144054	1200772	45	< 5	NA	AB	verified
NA	RGR8	failed	NA	1223614	1255961	36	0	1	(B)(C)	NA
MGD10	DGR10	successful	Yes	1273020	1309698	32	> 100	NA	AB	verified
NA	RGR9	failed	NA	1329945	1367729	36	0	0	NA	NA
NA	RGR10	failed	NA	1399532	1478908	70	0	0	NA	NA
NA	RGR11	failed	NA	1480695	1540442	51	< 5	1	A(B)(C)	NA
MGD11	DGR11	successful	Yes	1541577	1724834	167	NA	NA	D	verified
MGD12	DGR12	successful	Yes	1771365	1818026	43	< 5	NA	AB	verified
NA	RGR12	failed	NA	1821335	1850036	31	< 5	NA	A(B)	NA
NA	RGR13	failed	NA	1857673	1884890	29	> 100	NA	A(B)	NA
NA	RGR14	failed	NA	1884940	1910033	27	> 100	NA	AB(C)	NA
MGD13	DGR13	successful	Yes	1935996	1957779	23	0	1	(B)C	verified
MGD14	DGR14	successful	Yes	1960407	1987074	22	> 100	NA	AB	verified
NA	RGR15	failed	NA	1995538	2026373	30	0	1	(B)	NA
NA	RGR16	failed	NA	2033721	2068471	30	< 5	0	(A)(B)	failed
NA	RGR17	failed	NA	2098664	2195479	102	0	0	NA	NA
NA	RGR18	failed	NA	2215723	2238239	21	< 5	0	(A)(B)(C)	NA
NA	RGR19	failed	NA	2239452	2259450	15	0	0	NA	NA
MGD15	DGR15	successful	No	2287925	2308844	19	> 50	NA	AB(C)	verified
NA	RGR20	failed	NA	2324185	2358277	33	< 5	NA	(A)(B)(C)	NA
NA	RGR21	failed	NA	2399337	2418322	19	0	1	(B)(C)	NA
NA	RGR22	failed	NA	2447198	2493477	34	0	1	(B)(C)	NA
MGD16	DGR16	successful	Yes	2493883	2519587	24	> 50	NA	AB	verified
NA	RGR23	failed	NA	2560740	2591587	29	0	0	NA	NA
MGD17	DGR17	successful	Yes	2623088	2663563	38	< 5	NA	AB	verified
MGD18	DGR18	successful	Yes	2664390	2706074	47	8	NA	AB	verified
NA	RGR24	failed	NA	2707541	2737496	18	0	1	NA	NA
NA	RGR25	failed	NA	2739309	2776936	40	0	0	NA	NA
NA	RGR26	failed	NA	2883859	2909403	24	> 100	NA	(A)B(C)	NA
NA	RGR27	failed	NA	2918531	2938767	21	0	0	NA	NA
NA	RGR28	failed	NA	2995282	3021943	25	10	NA	(A)B	failed
NA	RGR29	failed	NA	3053283	3077896	22	0	1	(B)(C)	NA
NA	RGR30	failed	NA	3088981	3118052	24	< 5	NA	(A)(B)	NA

Table 3.2. Multiple-gene deletion validation results

Strain	Region	Deletion	MS growth	Start	End	CDS	GT colonies	GT puddle	PCR verifications	WGS
NA	RGR31	failed	NA	3184473	3216931	35	0	0	NA	NA
NA	RGR32	failed	NA	3241618	3264925	16	10	NA	(A)(B)(C)	NA
NA	RGR33	failed	NA	3321697	3371121	34	> 100	NA	(A)B(C)	NA
NA	RGR34	failed	NA	3372053	3404735	26	> 50	NA	AB(C)	NA
NA	RGR35	failed	No	3446141	3468761	24	> 100	NA	A(B)C	NA
MGD19	DGR19	successful	Yes	3506619	3530674	27	> 50	NA	ABC	NA
NA	RGR36	failed	NA	3570486	3598496	27	< 5	NA	A(B)	NA

Table 3.2. Multiple-gene deletion validation results (continued)
Multiple-gene deletion validation results, detailing growth in minimal succinate (MS) media, genome coordinates (green columns), number of annotated coding sequences (CDS), number of colonies from Golden Transformation (GT colonies), number of colonies from puddle transformation (GT puddle), PCR verifications for reactions illustrated in Figure 3.7 with failed verifications in parenthesis and whole genome sequencing verification.

Chapter 4: Mapping the adaptive landscapes of *Acinetobacter baylyi* large deletion mutants to open new paths to extreme genome streamlining

Suárez, G.A., Leonard S.P., Gangavarapu L.S., Barrick, J.E.

Research Contributions

Suárez, G.A. designed and performed the experiments. Gangavarapu L.S. contributed with DNA sample processing and validation. Suárez, G.A., Leonard S.P. and Barrick J.E. analyzed data. Suárez, G.A. and Barrick, J.E. wrote the manuscript.

Acknowledgements

We thank Daniel Deatherage for sequencing pipeline tips and technical support. We also thank Simon D'Alton, for providing space, organization and materials throughout evolution experiments. We acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for access to high-performance computing resources.

ABSTRACT

One goal of synthetic biology is to define the minimal components required for cellular life. One way to achieve this is to target portions of an organism's genome that are determined to be nonessential—via comparing genome sequences, constructing metabolic and regulatory networks, and testing the viability of single-gene knockouts—for deletion. Yet, systematic disruption of many genes often impairs robust cellular growth. On the road towards genome minimization, this loss of fitness can severely limit the number and depth of design trajectories that can be explored. Adaptive laboratory evolution is a common strategy for improving the fitness of bacterial strains subjected to stresses. We hypothesized that strains with large deletions would be able to rapidly acquire mutations that compensated for their fitness defects through altering their remaining genes. We evolved 11 multiple-gene deletion strains of *Acinetobacter baylyi* for ~300 generations in either rich or minimal media and found that one or two mutations can completely restore fitness in many cases. The recurring appearance of ribonuclease D (*rnd*) inactivating mutations in multiple evolved lines of different reduced genome strains suggests regulatory changes impacting global gene expression are a dominant adaptive response to generic losses in fitness due to widespread gene deletion. Specific deletion strains also had characteristic mutations in tRNA sequences, in the AbsR28 small regulatory RNA, or in the *csrA* gene that could compensate for processes that were compromised or unbalanced in these strains. Our results show that compensatory

evolution can overcome fitness defects that accumulate during engineered genome reduction to re-open paths to further genome minimization.

INTRODUCTION

Simplifying genomes by removing unnecessary and uncharacterized genes has the potential to make microbes into more efficient and predictable cellular factories [50]. The process of genome reduction can also help to better define the minimal requirements for cellular life [196]. Several efforts in different bacterial species have shown that large-scale genome streamlining can result in strains with other desirable traits, including improved predictability [197], greater stress tolerance [52,71], lowered autolysis [62], improved plasmid maintenance [55], and increased competence [59,63]. Notably, increased productivity (biomass yields) and improved expression of desired products (e.g., heterologous genes) have been achieved in both model and industrial organisms like *Escherichia coli* [56,66,68,198], *Bacillus subtilis* [62], *Corynebacterium glutamicum* [60] and *Pseudomonas putida* [64]. Often, the improvements are the direct result of removing components—such as proteases, nucleases, prophages and transposons—that interfere with expression levels or genetic stability related to the desired product [55,56,199]. Alternatively, the removal of non-essential gene clusters that encode particularly burdensome cellular functions (e.g., flagella or secondary metabolite production) may improve strain characteristics, for example, via upregulation of TCA cycle genes [62,64,66,200].

Despite these advances, minimal genome construction has often reached a dead end when strains accumulate growth defects that make further deletions effectively

impossible to construct [54,69,71,79,201]. In addition to losses in fitness due to deleting genes, genome reduction also carries the risk of inadvertently generating new sequence junctions that combine genetic components in ways that lead to growth defects. This was the case during the initial stages in the construction of the minimal *Mycoplasma mycoides* JCVI-syn1.0 genome, when colonies with severely impaired growth revealed that the designed deletion caused an unintended juxtaposition of a transcription terminator and an essential gene. However, colonies with faster-growing sectors began to appear within a 12-day incubation, demonstrating that it is possible for fitness problems that crop up during genome streamlining to be fixed by spontaneous mutations [54]. Evolution experiments also showed that *Salmonella* strains with large random deletions can recover fitness in < 100 generations [202]. Just as the cost of antibiotic resistance can often be mitigated by compensatory mutations [203], minimized genomes appear to still have access to a broad adaptive landscape that can compensate for fitness lost due to gene deletion.

Studies that examine the adaptive landscape of minimized genome strains are scarce and usually anecdotal. Although the most streamlined *E. coli* strain to date (Δ33a) has documented growth defects, an evolved variant (Δ33b) was later described as "fast-growing" [204]. It was reported that starvation-mediated mutagenesis and evolution improved growth and other phenotypic characteristics of the reduced-genome *E. coli* strain MDS42 [205]. Moreover, *Legionella pneumophila* with 18.5% of its genome deleted readily acquired point mutations that suppressed the deleterious effects of genome reduction [206]. Most recently, adaptive evolution of the streamlined *E. coli* strain MS56 revealed global transcriptional changes in gene expression due to mutations in the

housekeeping sigma factor (*rpoD*) that alter utilization of different promoters by RNA polymerase [86]. These types of global changes in transcription potentially make sense in the context of the increased expression of genes controlled by the stationary phase sigma factor (*rpoS*) that are observed in other reduced-genome *E. coli* strains (MDS42 and MDS69) [71]. Similarly, mutations in the transcriptional machinery (e.g., *rpoC*) are often observed as a means to optimize metabolism and improve growth under stressful conditions in many adaptive laboratory evolution (ALE) experiments with wild-type *E. coli* [207 208]. Thus, it appears that re-tuning the same global regulatory processes (e.g., transcription) may often be the most expeditious and generalizable way for evolution to improve the fitnesses of both wild-type genomes and less-fit reduced genomes, though the knob and degree of tuning may differ.

In some cases, loss of a specific set of genes may make it impossible for regulatory re-tuning to restore bacterial fitness to the same level, at least on a reasonable timescale. Because current evolutionary studies of minimal genomes have reported on fortuitous evolution experiments and/or examined only a single reduced genome, it is unknown how often compensatory evolution can or cannot restore fitness defects in minimized-genome strains. Much of our current understanding of how evolution compensates for the loss of different functions is derived from single-gene knockout studies. Some *E. coli* single-gene knockouts are able to restore their ability to grow in minimal medium after a mere ~145 generations of evolution [209]. The mutations behind this compensatory evolution are sometimes gene amplifications but they are mostly point mutations and IS elements [209]. Other studies have looked specifically at how

overexpression of a different gene can often compensate for depletion or complete loss of an essential gene [210]. Overexpression of one gene on a plasmid can often reverse an auxotrophy due to loss of another gene through secondary, promiscuous, or uncharacterized activities of a different gene even it they may not share any sequence homology with the deleted one [211]. Thus, there is widespread evidence that compensatory evolution can restore fitness after gene loss.

We hypothesize that purposefully employing compensatory evolution before combining deletions can allow one to overcome the fitness constraints that are encountered as more and more genes are removed during minimal genome construction. To begin to test this hypothesis, we selected 11 reduced-genome strains of *Acinetobacter baylyi* ADP1, each with 20-57 kb of the wild-type genome removed and performed short adaptive laboratory evolution experiments with each one in both rich and minimal media. Characterization of whether fitness was restored or not during these evolution experiments and of the compensatory mutations that appeared in each of these reduced-genome genetic backgrounds provides insights into the potential of utilizing evolution as a tool to further extreme genome reduction.

METHODS

Strains and culture conditions

Transposon-free *A. baylyi* ADP1-ISx is the parental strain of all deletion mutants in this study [211]. Liquid cultures were incubated at 30°C with orbital shaking at 200 r.p.m.

(1-inch diameter) in 18- by 150-mm glass test tubes in the Miller formulation of Lysogeny Broth (LB) or in minimal succinate (MS) medium [183]. Solid media included 1.5% (wt/vol) agar. Frozen stocks of strains and evolved populations were stored at –80°C with 15-20% (v/v) glycerol. Where appropriate, media were supplemented with 50 μg/ml kanamycin (Kan) or 200 μg/ml 3′-azido-2′,3′-dideoxythymidine (AZT).

All large deletion strains used in this study underwent AZT counter-selection to remove the selection cassette used for their construction from their genomes. AZT is toxic to cells harboring the *tdk* gene as part of the positive-negative selection cassette used to replace the deleted genes. The counter-selection procedure proceeded via a "rescue" Golden Transformation that combined the PCR products corresponding to the 5′ and 3′ flanks (~2 kb each) targeting the deletion region with plating the transformation mix on LB-AZT plates (**Chapter 3**). In every case, at least 2 colonies were screened via PCR for successful removal of the *tdk-kanR* cassette and further validated by testing for the expected inability to grow on LB-Kan agar.

Experimental Evolution

A total of 11 confirmed marker-less deletion strains and a wild type ADP1-ISx control were subjected to a ~300 generation adaptive laboratory evolution (ALE) experiment. To begin the experiment, each strain was plated onto LB and MS agar to obtain single colonies. After growth, 6 randomly picked clones of each strain from each type of agar were used to initiate independent 1 mL liquid cultures in the same medium to grow the Day 0 populations. On each of the following days of the evolution experiment

cultures were diluted 1:1000 into 5 mL fresh media, left to grow for 24 h and this process was repeated for a total of 30 cycles of serial dilution and regrowth (~300 generations). An exception was made for a very unfit deletion strain (MGD8). In this case an additional 24 h of growth was added when the cultured strain did not reach saturation in the first 24 h early in the experiment.

Day 0 cultures were archived as freezer stocks at −80°C. At the end of the experiment, all evolved populations were plated on LB or MS agar and a single large colony was selected in order to favor characterizing cells with compensatory mutations. Each of the Day 30 end-point clones was then grown overnight in 1 mL of the same medium as it was evolved in and stored as a freezer stock.

Fitness assays

Evolved population samples were competed against a GFP-tagged ADP1-ISx strain. This strain was constructed by inserting a *tdk-kanR* cassette with the T5 promoter from plasmid pIM1440 [117] in place of IS*1236*#6 [59]. Competition experiments began by reviving 5 µl of glycerol stocks into 5 mL LB cultures in test tubes. After 24 h of preconditioning growth, 5 µL of each of these cultures were used to inoculate a test tube containing 5 mL of LB. After 24 h, competitions for each population (n = 6) were initiated by mixing 3.5 µL of each evolved line with 1.5 µL of each GFP-ADP1 culture in 5 mL of LB; with the exception of competitions between progenitor GFP strain and evolved controls of ADP1-ISx, in which case 5 µL total of the two strains mixed in a 1:1 ratio was used. Immediately after mixing and after 24 h of growth, dilutions of these

initial and competed co-cultures were plated on LB agar. Lastly, counts of fluorescent versus non-fluorescent colonies were used to compute the relative fitness as a ratio of Malthusian parameters as previously described [212].

Whole genome re-sequencing

Genomic DNA was isolated from 1-2 mL cultures of Day 0 and Day 30 clones from freezer stocks in the same media in which they evolved using the PureLink Genomic DNA Mini Kit (Invitrogen) according to the manufacturer's instructions. DNA concentration was measured using a Qubit Fluorometer (ThermoFisher).

For a first set of 56 samples analyzed in this chapter, ~500 ng of DNA was fragmented with NEBNext dsDNA Fragmentase (New England Biolabs) for 27 minutes to generate ~300 bp fragments. The fragmented DNA was used as input to construct Illumina NGS-libraries using the LTP Library Preparation Kit (Kappa Biosystems) as described by the manufacturer except that reaction volumes were reduced by 50% and custom adapter sequences were used. 7 sets of adapters were used which had 6 bp of sequence immediately 3' of the standard Illumina TruSeq adapter sequence such that 8 individual samples could be sequenced using the same external barcode sequence. These libraries were sequenced on a NextSeq500 75 cycle pair-ended reaction at the University of Texas at Austin Genome Sequencing and Analysis Facility. Pools of 8 samples were demultiplexed using a custom python script based on the first 6 bases on each read.

Mutation analysis

Sequencing files associated with libraries were demultiplexed based on the 6-bp internal barcodes introduced during library prep using a custom Python script. Adapter sequences were removed from all reads using Trimmomatic (version 0.36) [213] before being evaluated for mutations. Trimmed reads were compared to the *A. baylyi* ADP1 reference genome [214] using *breseq* (version 0.33.2) [215] to confirm the presence of the expected deletions and identify additional mutations that accumulated in each strain during the course of the evolution experiments.to. Mutation figures were created using Circos [216] and R (R Core Team, 2019).

RESULTS AND DISCUSSION

Evolution Experiment

We conducted a 30-day (~300-generation) adaptive laboratory evolution (ALE) experiment on a collection of large deletion mutants by performing 1000-fold serial-dilution transfers in either LB or in minimal succinate (MS) media (**Fig. 4.1**). In total, the ALE experiment consisted of 144 populations: 11 large deletion mutants and their ADP1-ISx ancestor with 6 evolved lines in LB and 6 in MS. *A. baylyi* ADP1-ISx has had all transposable elements deleted from its genome [59] and each deletion strain has an additional large multiple-gene deletion of ~1% of its genome (**Fig. 4.2**). These strains are numbered according to their multiple-gene deletion (MGD) strain designation from **Chapter 3**. Strain designations are appended to the MGD number for independent ancestral isolates (labelled A1 and A2) and evolved isolates (labeled E1 through E6) followed by the medium in which they were evolved (LB or MS).

Fitness Evolution

All the deletion mutants in this adaptive evolution study were found to have impaired growth rates relative to the ADP1-ISx progenitor when cultured in rich medium in previous work (**Chapter 3**). Here, we performed more sensitive co-culture competition experiments, in the same medium (LB), to assay the changes in fitness before and after adaptive laboratory evolution. We began by removing the *tdk-kanR* cassette that was initially present in the genomes of each of the large deletion strains prior to the ALE experiment to exclude fitness costs (e.g., from production of the antibiotic-resistance gene) other than those generated by the deletions themselves.

The range of the growth defects in LB among the 11 deletion mutants varied from 38% to 89% (~61% average) of the growth rate relative to the progenitor strain (**Chapter 3**). We began to characterize fitness for parent and evolved strains. So far, we have performed fitness assays in LB with five of the deletion mutants (MGD3, MGD4, MGD6, MGD10 and MGD18). As expected, MGD4, MGD6, and MGD10 had reduced fitnesses at the start of the ALE experiment (**Fig. 4.3**). Unexpectedly, the MGD3 and MGD18 unevolved parent strains exhibited nearly the same fitness as ADP1-ISx (**Fig. 4.3**). We discovered through genome sequencing that this discrepancy was due to these two strains having already acquired compensatory mutations, probably during the marker removal step in constructing these strains (described below).

We judged whether there was a significant ($p \leq 0.01$) or marginally significant ($0.01 < p \leq 0.05$) increase in fitness in an evolved strain relative to its ancestor using two-tailed Student's *t*-tests. As mentioned, MGD3 had acquired mutations early on our

experiment, yet one of the evolved clones (E6) appears to have further restored fitness towards the ADP1-ISx baseline (**Fig. 4.2**). All six MGD4 LB-evolved clones have varying degrees of fitness recovery. In particular, four of MGD4 LB-evolved strains show significant—albeit partial—fitness recovery, while MGD4-E4LB shows no difference from wild type, and MGD4-E3LB exceeds ADP1-ISx fitness. The most remarkable case of fitness recovery is observed for MGD6, with stark contrast of fitness measurements between Day 1 and Day 30 clones. All but one of the MGD6 LB-evolved strains significantly recovered fitness, four partially and one (E6) with near ADP1-ISx fitness. One of the MGD10 LB-evolved clones (E2), marginally significantly increased its fitness ($p=0.045$), yet it remains much less fit than ADP1-ISx. No significant differences in fitness were observed for any of the evolved ADP1-ISx controls.

These results highlight that even when large deletions tend to be maladaptive, at least partial restoration of fitness is commonly achieved if propagation is allowed to attain compensatory mutations [202]. They are also generally consistent with the finding that the less-fit an initial mutation makes a strain, the greater its capacity to evolve increased fitness, since there are many more steep compensatory paths in the fitness landscape to re-ascend the same peak or access new peaks [217]. This is a type of positive epistasis that has also been seen with adaptive mutations that compensate for deletions, which are generally more beneficial in less-fit genetic backgrounds [218].

Genome Evolution Rates

We next examined the mutations that occurred during ALE for five of the deletion mutants (MGD3, MGD4, MGD6, MG10 and MGD18). Our dataset so far includes all six LB-evolved endpoint strains derived from each MGD ancestor, in addition to several of their MS-evolved counterparts. After 300 generations, we identified a total of 67 mutations in 42 sequenced LB and MS-evolved clones. The 67 mutations (9-15 per set of deletion mutants) can be classified into 6 categories, including large deletions, small indels, intergenic SNP, non-coding SNP, non-sense SNP and non-synonymous SNP (**Fig. 4.4**). Overall, we find ~1.6 mutations per evolved clone, with the number of mutations ranging from 1 to 3 (**Fig. 4.5**). Although a total of 23 genes/loci are mutated across all LB and MS-evolved clones, the majority (47/67) of the mutations happen within just six loci: *ACIAD0697*/lysP (AbsR28) (5 total), *ACIAD2521*/*ACIAD2522* (11 total), *ACIADtRNAVal_46* (6 total), *csrA* (5 total), *pgi* (4 total) and *rnd* (16 total) (**Fig. 4.6**). We observe a greater number of mutations in LB than in MS, with averages of 1.73 (52 mutations in 30 clones) and 1.25 (15 mutations in 12 clones) mutations per evolved clone, respectively; though this difference is not significant due to the small number of MS-evolved clones analyzed so far. Separating these out by ancestral MGD strain, there are 14 mutations in MGD3, 14 mutations in MGD4, 15 mutations in MGD6, 6 mutations in MGD10 and 15 mutations in MGD18. The smaller number of mutations observed in MGD10 correlates with no significant increases in fitness for all but one of its evolved clones. We found no mutations in the one evolved ADP1-ISx clone sequenced so far (E1). This is consistent with the negligible fitness increase in the six evolved ADP1-ISx

clones, but surprising because the first selective sweeps in ALE experiments usually happen within 300 or fewer generations. Since the remaining ADP1-ISx evolved clones did not get sequenced, it is possible these have some beneficial mutations. On the other hand, ADP1-ISx may evolve slowly due to a combination of its already high fitness in this environment (like MGD10) and its reduced mutation rate [59].

Specificity of Genomic Evolution

We next examined parallelism in the genes hit by mutations in different evolved strains and a previous 1000-generation ALE experiment with wild-type *A. baylyi* ADP1 in LB [19]. We found some recurring and overlapping mutations between the MGD strains and the prior ALE experiment (**Fig. 4.7**). The most frequent mutation across all evolved lines of different deletion strains corresponds to a 49-kb deletion of a prophage region. This deletion (coordinates ~2120339-2,169,692) was also previously observed in wild-type ADP1. Its appearance across so many genetic backgrounds makes it an example of a universally beneficial mutation (i.e., one that is not dependent on the gene deletion background). Previously observed intergenic *ACIAD2521-ACIAD2522* mutations (immediately upstream *ACIAD2521*) and mutations in the *pgi* and *per* genes were prevalent across different deletion strains and also fall into the same category. In contrast, recurrent mutation of *rnd* (RNase D) and *csrA* happened in multiple deletion strains but not in the evolved ADP1-ISx control we have sequenced or in the previous ALE experiment with ADP1. These mutations may be a generic adaptive response to compensating for some shared characteristic of deletion strains with reduced fitness (e.g.,

growth rate). Other common mutations are specific to one MGD strain background such as mutations affecting the AbsR28, which are common in and only appear in the MGD6 background.

It has been observed that strains with deletions in the same functional module tend to adapt similarly [218]. Other studies have repeated this observation that most compensatory (suppressor) mutations following gene deletion happen within the same functional network but also noted that they are not necessarily limited to it [219 220]. Previous studies have also found that compensatory mutations in single-gene knockouts often happen in the exact same remaining gene (parallel evolution) [221 220]. We also observe some targets for adaptive mutations that seem to be in common to all large deletion strains and others that are specific to MGD strains in which certain pathways or genes are removed that compromise fitness or constrain adaptive pathways.

Universally beneficial mutations in extracellular polysaccharide production

During our AE experiment, several of the LB-evolved ADP1-ISx and large deletion mutants acquired mutations impacting perosamine synthetase (*per*; *ACIAD0095*) or glucose-6-phosphate isomerase (*pgi*; *ACIAD0101*) genes (**Fig. 4.6**). Specifically, of the sequenced deletion mutants, 3/5 (MGD 3, MGD4 and MGD18) harbored *pgi* mutations in one or multiple evolved lines. Notably, mutations impacting the *pgi* gene are prevalent (3/6) in LB-evolved MGD18 (E2, E5 and E6), while only found once in MS-evolved MGD3 (E1), and also once in MS-evolved MGD4 (E3). Similar *pgi* and *per* mutations have been observed previously in ALE experiments performed in our lab and presumed

to be beneficial due to their high frequency [19]. The *pgi* gene encodes phosphoglucose isomerase, which plays a key role in glycolysis and gluconeogenesis pathways, catalyzing the interconversion of glucose-6-phosphate and fructose-6-phosphate. Both *per* and *pgi* are known to be involved in the biosynthetic pathway of a secreted emulsifier in the closely related species *Acinetobacter lwoffii* RAG-1 [222 223]. Specifically, they belong to a large multi-gene cluster known as *wee* — also known as the *capsule gene cluster* — which is required for biosynthesis of emulsan polysaccharides [224 222 223]. In wild-type ADP1, inactivation of the *pgi* gene was determined to be involved in cellular aggregation and causes cells to settle at the bottom of liquid cultures [19], a phenotype we also observed in LB-evolved MGD18. Additional mutations and deletions (Δ195 bp in *ACIAD0098* of a MGD3 line and Δ11,507 bp covering genes *ACIAD0093–ACIAD0103* of a MGD18 line) are found throughout the *wee* gene cluster, including mutations in *ACIAD0085* – a glycosyl transferase – and in *galU* (*ACIAD0099*) (**Fig. 4.5**). Similarly, large IS-mediated deletions that span these same genes were observed in prior ADP1 ALE experiments [19]. Overall, due to the frequency and appearance of *wee* gene cluster mutations across multiple evolved lines regardless of the genomic context, these can be considered universally beneficial. They are good candidates for mutations that could be incorporated into any future reduced-genome strains to give them greater fitness.

Universally beneficial mutations upstream of *ACIAD2521*

In addition, several point mutations and a small deletion (deletion of one copy of TAAT repeat) arose upstream (98-146 bp) of the start codon of *ACIAD2521*, a gene of

unknown function (**Fig. 4.8**). Of these, the exact same base substitutions at 127 bp upstream of the start codon of this putative gene were also noted multiple times in the previous wild-type ADP1 ALE experiment [19]. In that study, it was suspected that these variants emerged as a polymorphism in the parent ADP1 culture used to initiate the AE experiment due to their high frequency. Given the independent origins of our deletion mutants, our data suggest instead that mutations in this region increase in frequency quickly. Note that even though the small TAAT deletion seems to have been carried from the MGD3 parent (it is in unevolved MGD3 and all of its evolved clones), suggesting that it occurred during deletion construction in this lineage, this exact same mutation was only prevalent (9/12) in MGD4 and was seen just once in MGD6-derived strains. Also, while MGD10 has recurrent point mutations at 98 bp upstream of the start codon of *ACIAD2521*, a different point mutation (146 bp upstream of start codon) shows up exclusively in one of the MGD4 evolved lines. All of these mutations—covering a 49-bp region—likely affect expression of the *ACIAD2521* gene by modulating its promoter. The lack of genetic background context-specificity of these mutations suggests they benefit both genome-reduced and wild-type genomes.

RNAse D mutations specific to MGD strains

A total of 16 *rnd* mutations are observed (in both LB and MS media), yet these are only prevalent in MGD3, MGD4 and MGD6. Of these *rnd* mutations, we note a recurring deletion (62 bp) and several premature mutations to stop codons near the end of

the reading frame (**Fig. 4.9**). The spectrum of the *rnd* mutations suggests that they reduce or eliminate *rnd* activity and that this change is somehow beneficial to cells.

RNAse D is an exoribonuclease that participates in the 3' maturation of tRNA, 5S rRNA, and several other small-structured (tRNA-like) RNAs [225] [226]. It is most active on tRNA molecules containing additional residues following the mature 3′ terminus or molecules lacking all or part of the −CCA sequence. It also has high specificity to denatured and damaged tRNA substrates. In *E. coli,* strains devoid of RNase D grow normally and show no defect in tRNA processing [227], while overexpression has been determined deleterious [228]. It is only essential for viability when other nucleases (RNAse II, BN, T and PH) are eliminated, suggesting a role as a backup enzyme when primary nucleases are missing [229]. This gene has a conserved DEDD catalytic motif with a fold that is similar to the Klenow fragment exonuclease domain, and this motif along with two putative nucleic acid domains come together to form a ring-shaped structure important for substrate specificity [226].

Previously, *rnd* mutations were suspected as the possible cause behind the observed gene expression changes critical for pathogenesis, by way of transcriptome remodeling during infection [230]. This was determined by the observation that *rnd* gene inactivation (IS-element mediated) were prevalent in isolates from patients infected with *Acinetobacter baumanii,* a close relative of *A. baylyi*. In our study, ADP1 is responding similarly (*rnd* inactivations), yet in a very different environment (LB), and in various multi-gene deletion contexts. The relationship between multi-gene deletions and pathogenesis, although disparate, could be linked in that both may be related to the roles

of *rnd* in post-transcriptional regulation. As a global transcriptome modulator, *rnd* inactivation or modulation could cause large global changes in gene expression capable of leading to large fitness benefits in a single mutational step. Reduced-genome variants of other bacterial species have shown to have greatly perturbed gene expression patterns, including higher expression of stress related genes [71] [53]. Future transcriptome analyses of the effects of the *rnd* mutations reconstructed in wild-type and reduced genome ADP1 variants could elucidate how it is affecting global gene expression in a beneficial manner and to what extent these benefits may depend on some generic feature of deletion strains (e.g., reduced growth rate).

Mutations in small RNA AbsR28 specific to MGD6 strains

Five of ten sequenced MGD6 evolved lines acquired intergenic point mutations (all different) between genes *ACIAD0697* and *ACIAD0700* (*lysP*). All of these mutations happen downstream of both these genes, rather than upstream in possible promoter regions, so it was unclear how they could affect these genes. This led us to looked for the presence of unannotated genes in this putative intergenic region. By querying the Rfam database, we identified an *A. baylyi* match to a small RNA, AbsR28 (RF02606), that had been experimentally validated in *Acinetobacter baumanii* [191] [231] (**Fig. 4.10**). One predicted target of AbsR28 – from using the IntaRNA tool [232] – maps to *ACIAD2197*, a gene commonly deleted as part of the known ~ 49 kb prophage region. It may be possible that these sRNA mutations may regulate phage functions or prevent prophage-related

excision from the genome, since the aforementioned prophage region is deleted often (9 cases) in other MGD strain backgrounds but never in MGD6.

Carbon storage regulator *csrA* mutations

Several evolved lines (4/6 LB-evolved) of MGD4 acquired *csrA* non-synonymous point mutations, and *csrA* was also mutated in one of the MGD6 evolved lines (**Fig. 4.11**). The *csrA* reading frame was altered during IS*1236*#4 removal during construction of the ADP1-ISx strain that is ancestral to all MGD strains. Deletion of this IS-element removed 13 of the normal amino acids from the C-terminus of CsrA and added translation of 26 new amino acids to this protein before the new stop codon is reached. It is unknown how these changes to the C-terminal end of CsrA affect its function.

The carbon storage regulator CsrA is a stress and global response gene known to mediate gene expression changes during shifts from rapid growth to stress survival [233 234]. The appearance of mutations in *csrA* in our study could also be consistent with the induction of a general stress response in reduced genomes [71]–provided that altering the function of the perturbed protein produced in the ADP1-ISx strain changes this activity, rather than just restoring it to its wild-type function. As a key translational regulator, CsrA binds its target mRNAs to regulate translation initiation and/or mRNA stability. It is possible that mutations in *csrA* result in the repression of phosphoglucose isomerase (*pgi*), since a fully functional *csrA* gene would be expected to positively regulate the activity of *pgi*. Thus, loss of *csrA* could repress *pgi* to phenocopy beneficial mutations in *pgi* found in other evolved strains [19,233]. The lack of simultaneous *pgi* and *csrA* mutations

in any of our evolved strains is consistent with this possibility, though it needs to be tested directed by reconstructing and testing strains with combinations of these mutations. Since *csrA* has been shown affect the expression of hundreds of genes in *E. coli*, mutations in this gene may be a further indication that deletion mutants are prone to undergo global transcriptome remodeling to restore growth rates [235].

CONCLUSION

We sought to elucidate the genetic basis of mutations that compensate for or otherwise reduce the fitness defects of a panel of *Acinetobacter baylyi* ADP1-ISx strains with different multiple-gene deletions. To do so, we sequenced ancestral and end-point clones from a ~300 generation laboratory evolution experiment. Notably, multiple mutations in the *rnd* and *csrA* genes spontaneously arose in deletion mutants during ALE, evidence that genome-reduced strains likely undergo rapid global transcriptional changes to counter metabolic imbalances due to the loss of multiple genes. In addition, certain deletion mutants gained compensatory mutations that were unique to the specific genes that they were missing, possibly as mechanisms to restore fitness defects from metabolic or regulatory imbalances caused by the deletions. Other universally beneficial mutations were observed to restore fitness in several of the deletion mutants and also in the control strain that does not have a deletion. In most cases characterized so far, near or complete fitness restoration was observed. Lastly, our study reveals that, in the case of the multi-gene deletions we have investigated so far, adaptive mutations do not seem to be

restricted to functional modules that are obviously related to the deleted genes in each deletion mutant.

In sum, this work builds a framework to further our understanding of how reduced-genome variants can evolve to restore their fitness defects. If a thorough landscape of compensatory mechanisms can be mapped, it paves the way to superimpose multiple deletions and compensatory mutations to possibly achieve genome streamlining without compromising fitness, at least to the degree seen in most strains with severely reduced genomes that have been created to date. The sets of evolved *A. baylyi* ADP1-ISx multiple-gene deletion strains discovered here may also create opportunities for overcoming fitness constraints to explore even more extreme genome streamlining design trajectories than is currently possible. These insights will further the synthetic biology vision of engineering minimal genomes for use in industrial and research settings.

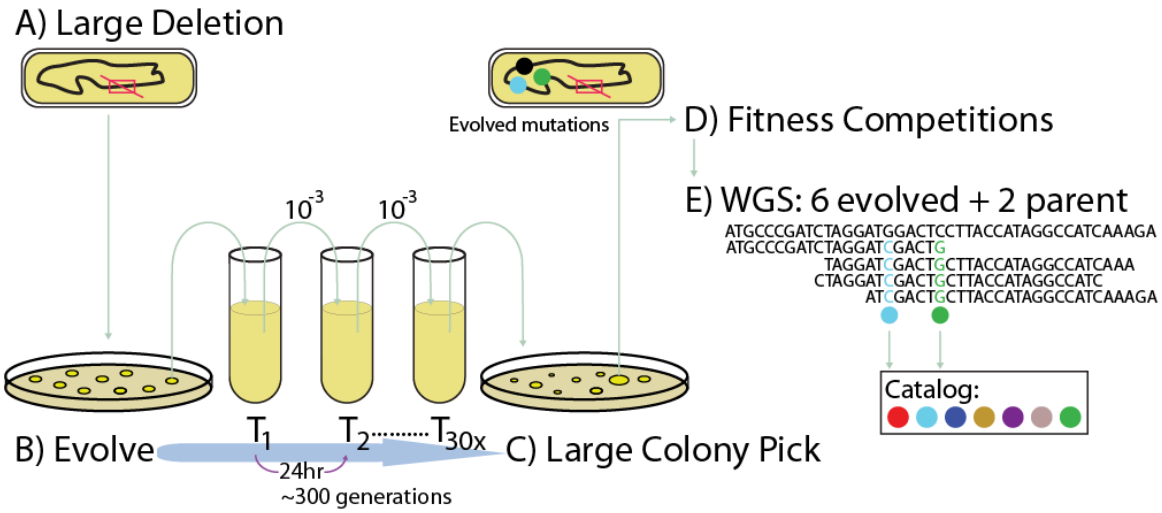

Figure 4.1. Adaptive evolution experiment.

A set of 11 large deletion mutants were subjected to a ~300 generation adaptive laboratory evolution. Starting from single colonies twelve populations of each of the deletion mutants were passaged with 1000-fold dilutions in 5 mL of either rich medium (LB) or minimal succinate medium (MS). A single large-colony clone was isolated from each evolved population after 30 transfers. The fitnesses of these clones were characterized using competition assays and mutations they evolved were determined by whole-genome re-sequencing.

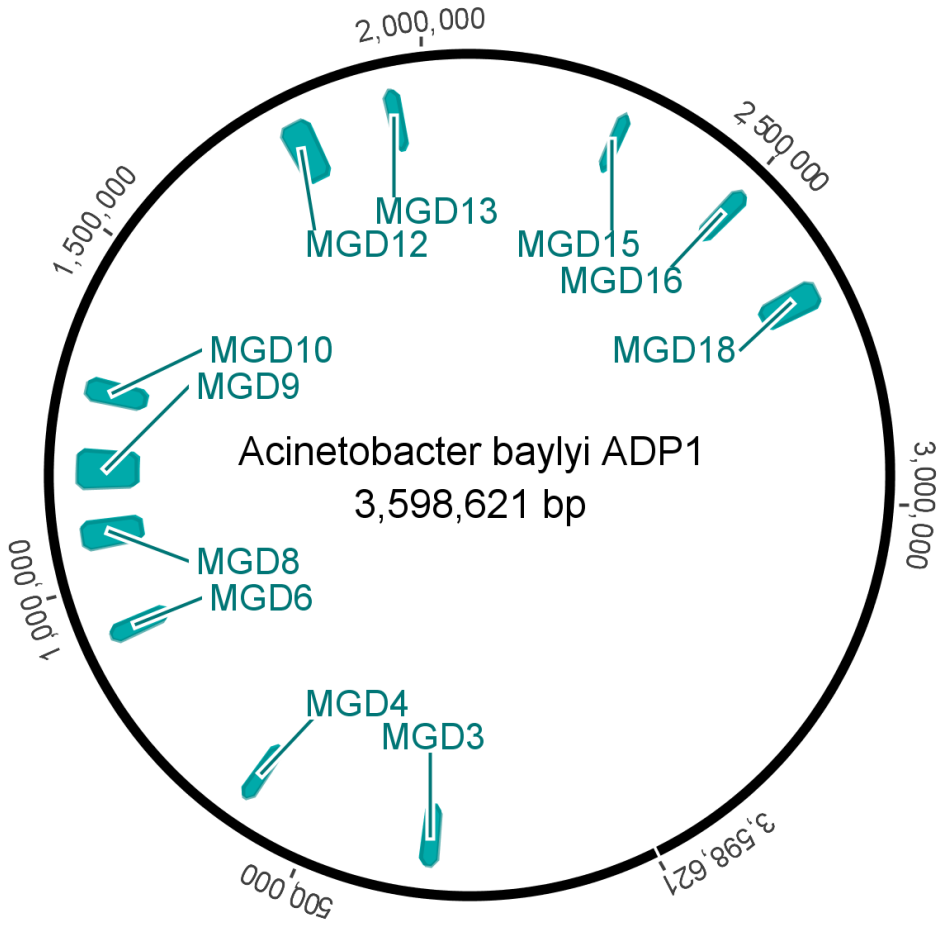

Figure 4.2. Large deletion mutants selected for the AE experiment.

Regions deleted in each of 11 large deletions mutants used in this study are shown on the 3.6 Mb *A. baylyi* ADP1. These are a subset of a previously generated ADP1 large deletion collection (**Chapter 3**). The deletions in the 11 MGD strains selected range in size from 20.9-56.7 Kb and remove 19-46 dispensable genes.

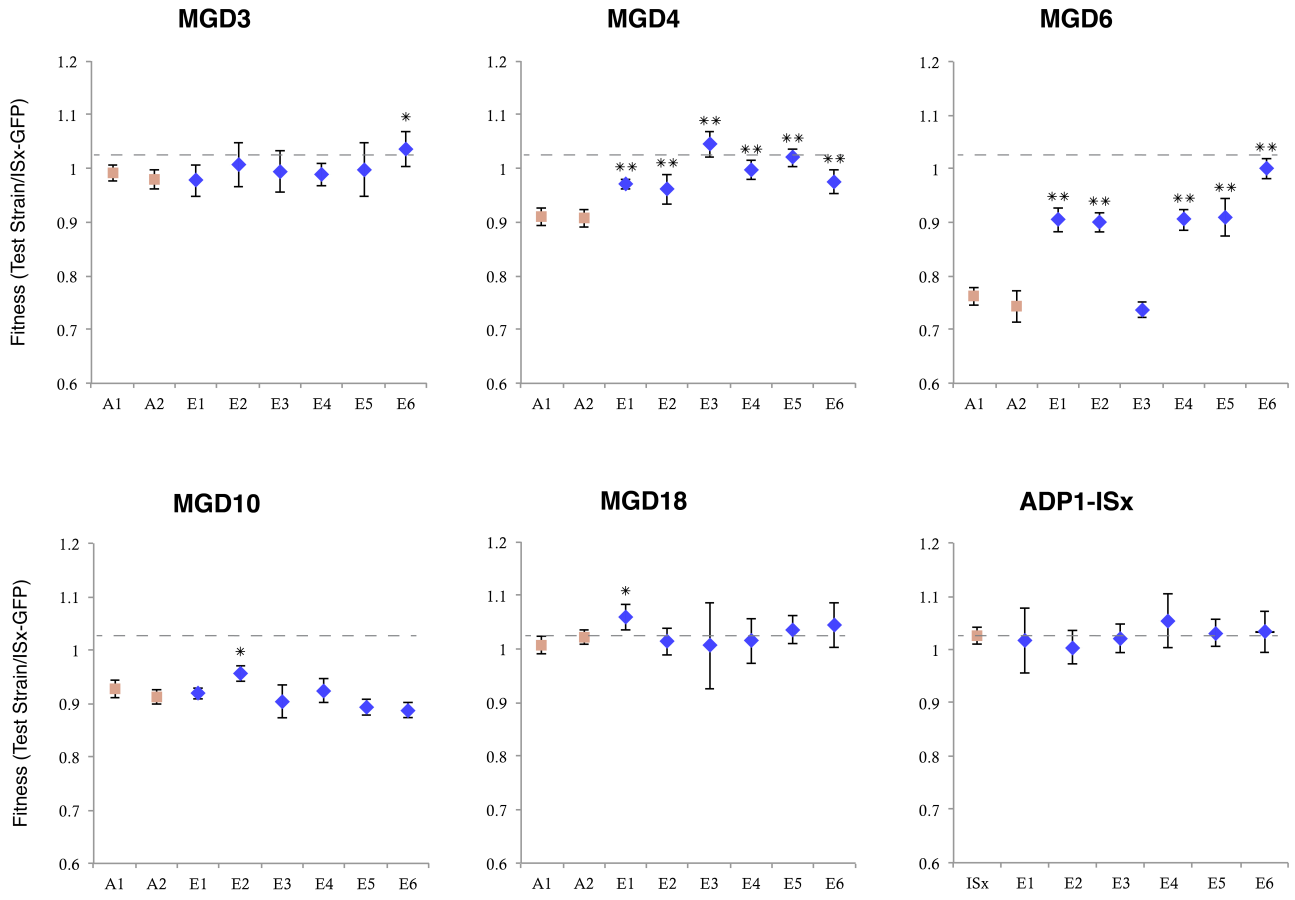

Figure 4.3. Relative fitnesses of evolved and ancestral MGD strains.

Five of the large deletion mutants tested, two control unevolved original progenitors (brown) and six evolved lines (blue) were competed against a GFP-marked wild-type ADP1-ISx strain in LB. Significance is determined by two tailed student's t-test with Bonferroni correction and indicated by one (p-value <0.01) or two (p-value <0.001) asterisks over the corresponding data point.

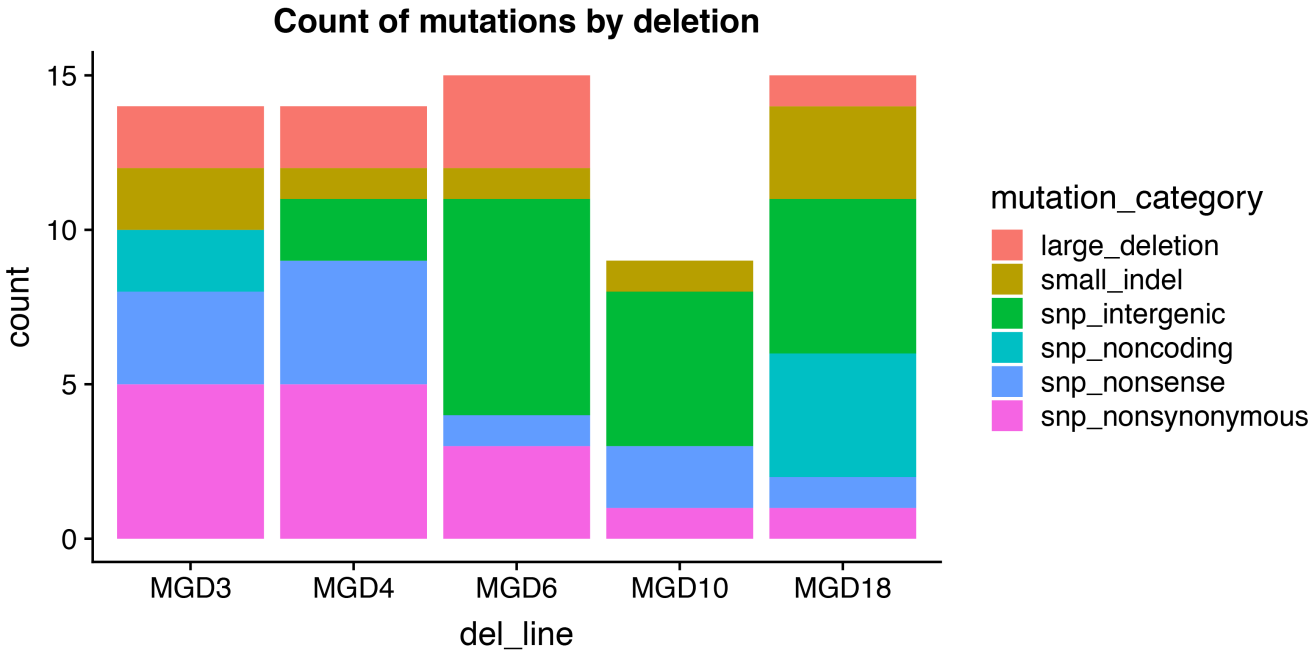

Figure 4.4. Mutation categories by deletion mutant.

Compensatory mutations in adaptive laboratory evolution of large deletion mutants for five deletion mutants (6-12 replicate lines for each in LB and minimal media) after experimental adaptive evolution. The sum total of mutation types found in any of the deletion mutants are displayed and classified into six categories represented by their respective color: large deletions, small indels, intergenic SNP, non-coding SNP, non-sense SNP and non-synonymous SNP.

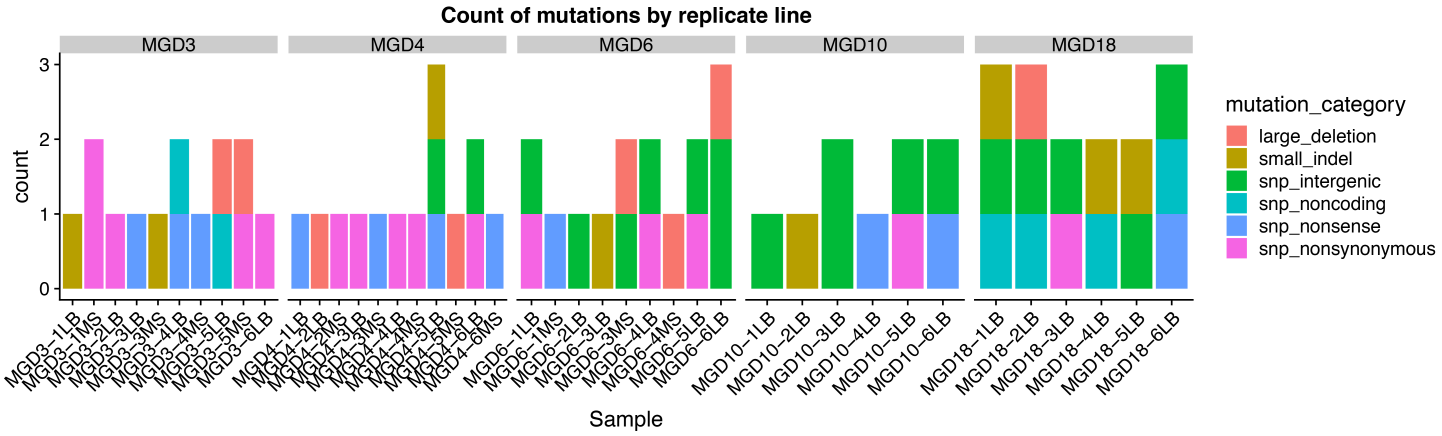

Figure 4.5. Distribution of mutation categories in each of the evolved large deletion mutants.

The mutation categories in each of the deletion mutants are analyzed for 42 of the LB-evolved and MS-evolved mutants. Each deletion mutant acquired 1-3 mutations.

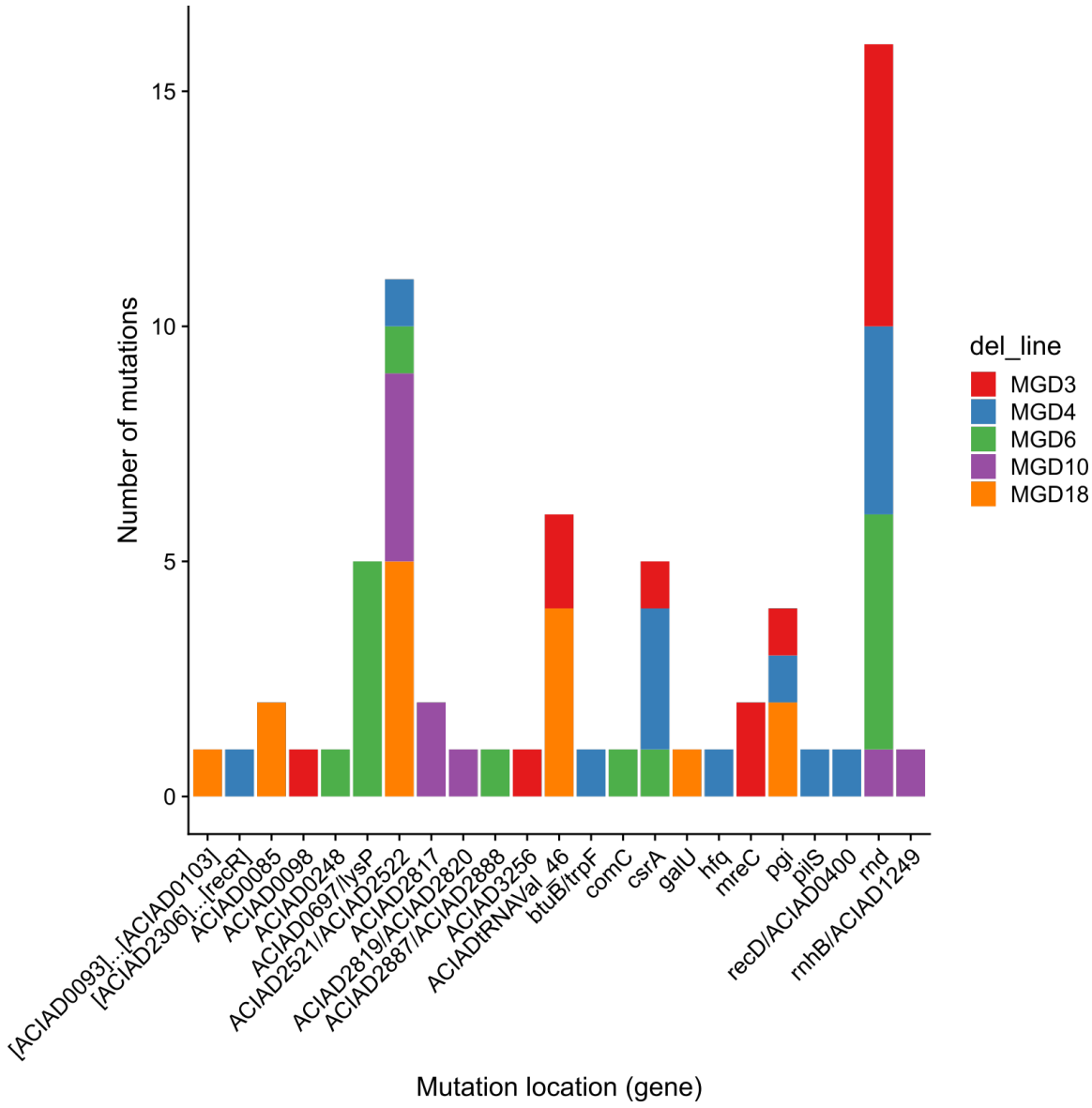

Figure 4.6. Mutation counts per gene/loci per deletion mutant.

There were a total 23 mutated genes/loci found on any of 5 LB or MS-evolved deletion mutants. Of a total of 67 mutations, 47 happen within the six most frequently mutated genes/loci: *ACIAD0697*/*lysP* (AbsR28) (5 total), *ACIAD2521*/*ACIAD2522* (11 total), ACIADtRNAVal_46 (6 total), *csrA* (5 total), *pgi* (4 total) and *rnd* (16 total).

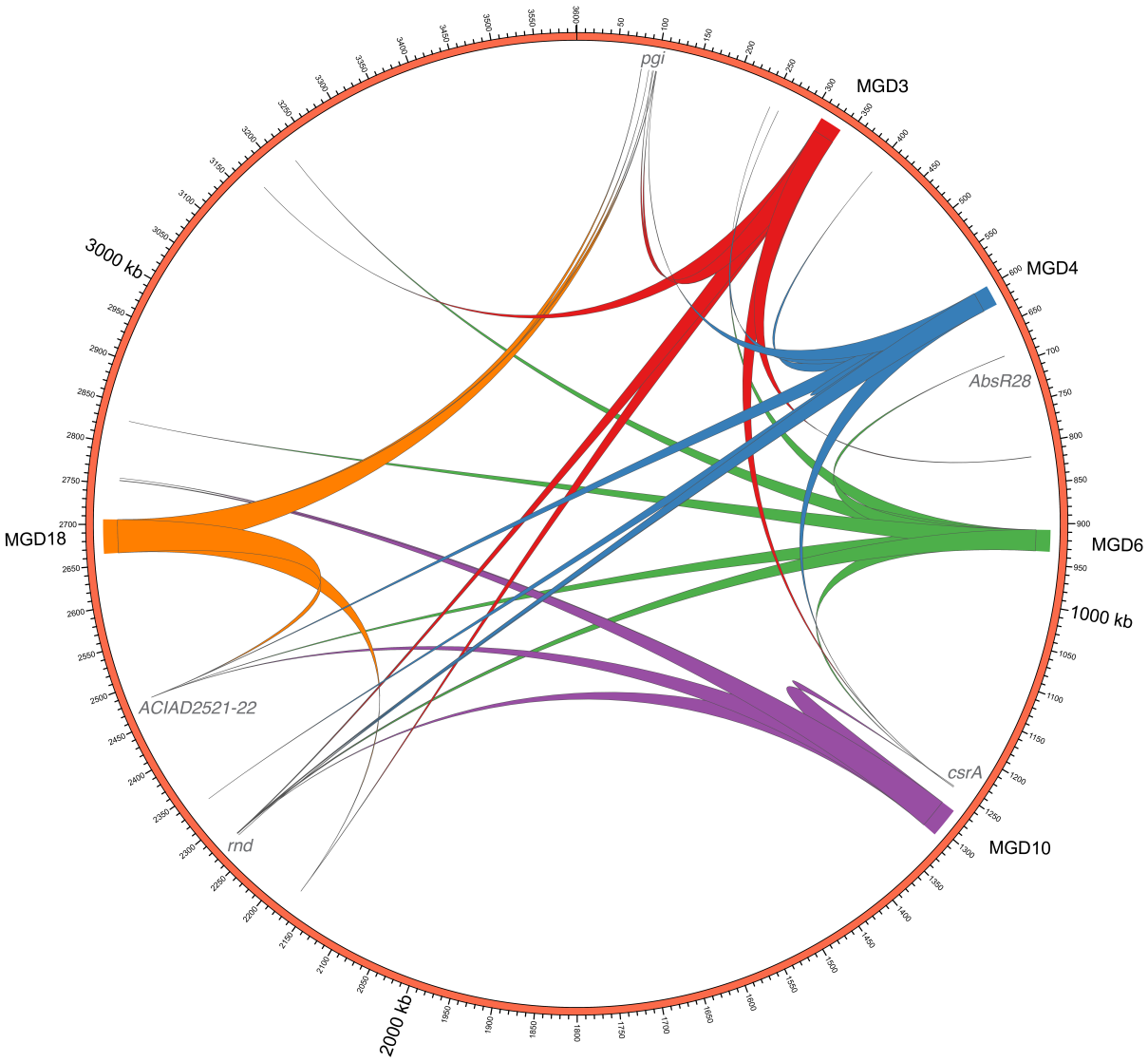

Figure 4.7. ADP1 chromosomal distribution of compensatory mutations of deletion mutants.

View of all LB-evolved and MS-evolved compensatory mutations identified by whole genome sequencing throughout the 3.6 Mb chromosome of *Acinetobacter baylyi* ADP1, derived from each of the 5 large deletion mutants submitted to the ~300 generation adaptive laboratory evolution experiment. Sites of parallel evolution, including the *rnd* gene, small RNA mutations and the *csrA* gene, are highlighted.

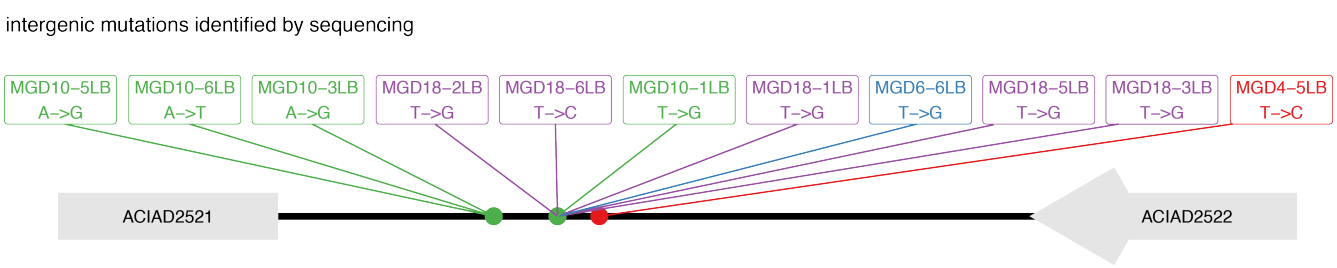

Figure 4.8. Mapped intergenic ACIAD2521-22 mutations.

A total of 11 point mutations, between *ACIAD2521* and *ACIAD2522*, are seen across four of the large deletion mutants: MGD4 (red), MGD10 (green), MGD6 (blue) and MGD18 (purple). The mutations span a region within 98-146 bp upstream of *ACIAD2521*, a gene of unknown function. No other smaller proteins are found within this region (NCBI ORF finder).

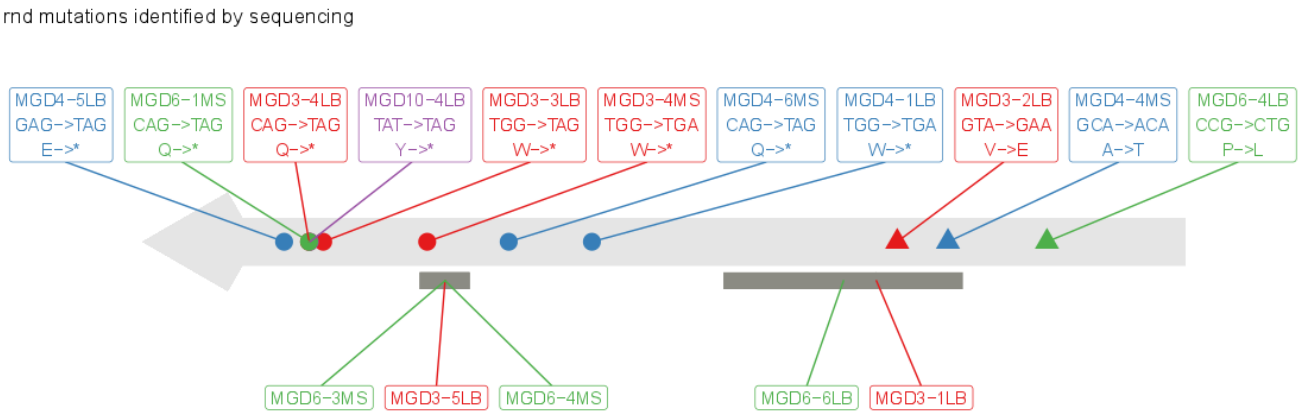

Figure 4.9. Mapped *rnd* mutations.

Location and description of mutations in the *rnd* gene. Label/symbol colors represent any of the 4 evolved large deletion mutants where these were acquired: MGD3 (red; circle), MGD4 (blue; circle), MGD6 (green; triangle), MGD10 (purple; circle). MGD18 did not acquire *Rnd* mutations, thus is excluded. Mutation types included non-synonymous point mutations, stop codon mutations (TAG or TGA) or deletions (represented by gray bars at the affected locations).

csrA mutations identified by sequencing

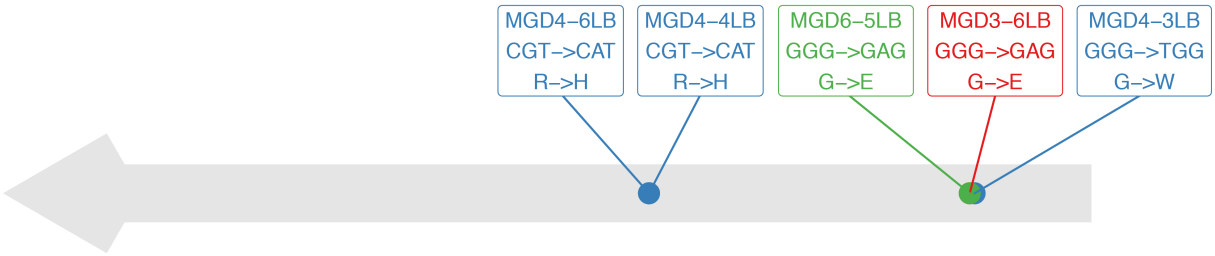

Figure 4.10. Mapped *csrA* mutations.

A total of 5 *csrA* mutations are seen across three of the large deletion mutants: MGD3 (red; circle), MGD4 (blue; circle) and MGD6 (green; circle). It is possible these mutations are evolved protein adjustments resulting from deletion of last 13 amino acids during ADP1-ISx construction.

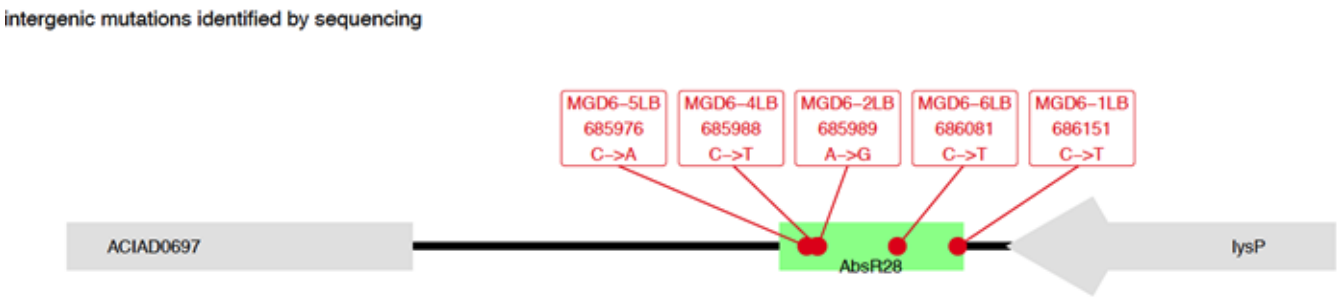

Figure 4.11. Mapped mutations of AbsR28 small RNA.

Multiple lines of LB-evolved MGD6 acquired mutations downstream the *lysP* gene. This led to the identification of AbsR28 small RNA, depicted with the blue green bar (coordinates 685,945-686,157 in the ADP1 genome).

Chapter 5: Conclusions and future directions

Engineering greater *A. baylyi* genome stability: remaining challenges

Several other factors that impact the genome stability of *A. baylyi* ADP1 as a platform organism remain to be explored and addressed in future work. With the IS-element free ADP1-ISx strain created in this work, a significant portion of the mutational load was successfully reduced. Yet, there remain selfish genetic elements lurking in its genome that could be removed in future work, and there is still the potential to reduce mutagenic processes affecting the fidelity of DNA repair and replication to further improve its genome stability [236 237].

Importantly, clearing phage and phage-related proteins from the ADP1 genome remains to be addressed. These additional regions of likely genomic instability contain a large number of genes classified as non-essential (58 total; ~20 predicted to be phage-related). One prophage region is located in an unstable part of the genome flanked by repeat sequences. It is often spontaneously lost during evolution experiments. Thus, it poses little problem for removal. However, another prophage, which can become activated during growth of ADP1 in the laboratory to produce the competence-reducing *Acinetobacter* phage (CRAϕ), is more troublesome [19 238].

CRAϕ is a filamentous phage that appears to utilize the competence pilus as a receptor, as knockouts of competence genes provide immunity. Stochastic activation of phage production in some cells in *A. baylyi* ADP1 populations leads to strong selection for mutations that knock out competence during evolution experiments. Deletion of this

prophage would therefore stabilize maintenance of this most valuable trait for ADP1 genome engineering. However, we have been unable to inactivate the CRAϕ prophage after numerous attempts at replacing or inserting selection cassettes within its sequence. In cases where it is successfully removed from the chromosome, it persists in its episome form within cells, indicating that it may encode a toxin-antitoxin module or some essential function that leads to inviability when it is removed. In the future, it is important to remove this and the other prophages from the ADP1-ISx strain to achieve a fully clean-genome strain with even greater genomic stability.

One can also evolve or engineer bacteria that have lower mutation rates by altering DNA replication and repair processes [237]. For example, specific mutations in *E. coli* DNA polymerases (e.g., *polA* and *polB*) can lead to lower rates of replication errors. This is presumably the result of increasing nucleotide selectivity or reducing extension of mispaired primer termini in the case of PolA mutants [239], whereas knockout of the error-prone polymerase PolB allows other higher-fidelity processes to be used for repair [199]. Additionally, certain aspects of the DNA damage response can be engineered to optimize a cell's capacity for preventing mutagenic damage. For example, *E. coli* with inactivating mutations in enzymes that lead to the production of reactive oxygen species have been shown to have lower plasmid mutation rates [240]. Such mutations could potentially be added to ADP1-ISx to further increase the fidelity of genome replication, and this—in principle—would enable more stable maintenance of burdensome and complex biosynthetic designs over many generations of growth.

In turn, it may be possible to overexpress DNA repair and other genes to reduce mutation rates. Candidate genes whose expression levels have been linked to mutation rates in *E. coli* include RNase H (*rnhA*), for its role in clearing away RNA-DNA hybrids, and single-stranded binding protein (SSB), for its role in protecting single-stranded DNA during replication and recombination. Controlled overexpression of RNAse H reduced amplification and point mutations by 60% on *E. coli* without reducing growth or cell viability [241]. These types of genes targeted for upregulation could be tested by placing copies of them on a plasmid into ADP1-ISx before deciding whether it is helpful to engineer the native gene for overexpression to stabilize the genome.

Development of novel engineering tools for ADP1 and other strain improvements

Our demonstration that the native type I-Fa CRISPR-Cas locus is active in *A. baylyi* ADP1 and the method we developed to retarget it using Golden Transformation open new opportunities. Tools based on CRISPR-Cas systems have revolutionized the way we perform targeted genome manipulations [242]. Yet the large diversity of CRISPR-Cas family types still provides an ample opportunity to explore their multiple roles and mechanisms in bacterial adaptive immunity [193], and characterization of new varieties of these systems may foment the development of new and improved tools [243] [244]. In this work, we have shown that it is possible to reprogram the endogenous Cas operon and CRISPR array in *A. baylyi*. This knowledge opens up the possibility of repurposing it – via *cas3* knockout - for transcriptional regulation, to further expand the *A. baylyi* toolkit for genetic control [195].

Other directed evolution approaches to minimizing genomes

Although deletions happen spontaneously during experimental evolution, the rates are generally too slow (0.05 bp per chromosome per generation) to be of practical use for genome reduction applications [202]. Tampering with DNA repair pathways (e.g., impairing methyl-directed DNA mismatch repair) can increase these rates ~50-fold [202], but the rate is still too slow to be of use. Furthermore, this strategy will also lead to the accumulation of deleterious mutations [245]. Selection strategies can sometimes isolate spontaneous mutants with very large deletions (> 200 kb) in their chromosomes. These can be exploited to generate a new variety of reduced genomes—which may never have been planned because of the sets of genes they include—to expand streamlining potential even further [202].

Along these lines, some of our work that did not bear fruit in the end involved testing a Tn-Seq-like [246] combinatorial deletion methodology that we called rHAT for random Homology-Aided Targeting). rHAT aimed to exploit the natural competence of *Acinetobacter baylyi* ADP1 by feeding it an antibiotic selection marker randomly re-ligated to ~1-kb fragments of genomic DNA to increase the rate of deletions and rearrangements in its genome. This method allowed for antibiotic selection of deletion mutants, followed by characterization at the site of cassette integration by PCR and sequencing. Of the 18 clones characterized, 10 (56%) had evidence of the same types of duplications of large genome regions that we often observed when trying to delete regions of the ADP1-ISx genome that contained essential genes (**Chapter 4**), and the putatively successful deletions were relatively small (**Table 5.1**).

We also tried to add the rHAT DNA to evolution experiments (without antibiotic selection) to uncover more-fit deletion mutants to generate a more comprehensive list of large sets of genes or dispensable regions for ADP1. To our disappointment, after 30 daily transfers (300 generations of evolution) with the randomly religated DNA we found –via whole genome sequencing– only one large deletion (coordinates 3,240,894 - 3,262,701) among the five replicate LB-evolved lines. This could have been the result of (1) low mutation rates, if there was a low frequency of integrations generating deletions were rare, (2) or selection effects, if most deletions were deleterious to fitness or at least less-beneficial than other mutations that were possible and occurred at a high frequency. Future work could explore combining these approaches with the CRISPR-lock strategy (**see Chapter 3**) to select for deletions in a way that prevents amplification. These undirected, combinatorial methods are interesting because they have the potential to reveal unexpected regions of genome dispensability in *Acinetobacter baylyi*.

Next steps toward minimal-genome *Acinetobacter baylyi* strains

While the research described here did not focus on the creation of multi-deletion minimal-genome strains, it did identify a path of inquiry that leads in that direction. In the future, solutions provided by allowing compensatory evolution after deletion of multiple genes can be integrated in genome streamlining designs for greater flexibility. The beneficial mutations identified in this work can be engineered into a multi-deletion strain to overcome fitness constraints that may arise due to genome reduction steps. Relocation of one or a few essential or quasi-essential genes from larger dispensable regions to

another part of the genome can also potentially be used to further minimize genomes [54] [192].

Additionally, using RNA-seq to track gene expression levels in reduced-genome strains during growth would provide an ampler view of how genome reduction is affecting cellular systems. When combined with measuring baseline wild-type expression levels, RNA-seq could also be used to test the effects of fitness-compensating mutations that appear to direct global transcriptome changes, including *rnd* and *csrA,* on stress and growth-related gene expression to see which genes they control in ADP1. Once regulators or their important downstream targets are identified, the effects on the fitness of reduced genome platforms of purposefully altering their expression can be tested. Together, these strategies can potentially allow a researcher to explore otherwise unattainable genome reduction paths and open up opportunities for more ambitious approaches to get closer to a fully defined free-living minimal cell.

Altogether, this work has provided a number of solutions and insights into improving *Acinetobacter baylyi* ADP1 as a platform for biological engineering. These developments and foundations will empower more ambitious genome editing projects that can be used to engineer cells that produce renewable fuels and green chemicals or that sense and remediate environmental contaminants, among many other possibilities.

Strain	Start	End	Length	Frequency
ISx-rHat1	150,661	156,431	5,771	1
ISx-rHat2	515,279	521,312	6,034	1
ISx-rHat3	543,651	549,215	5,565	2
ISx-rHat4	1,093,164	1,099,125	5,962	1
ISx-rHat5	2,267,738	2,272,388	4,651	1
ISx-rHat6	3,172,273	3,181,499	9,227	1
ISx-rHat7	3,533,237	3,535,068	1,832	1

Table 5.1. ADP1-ISx deletion mutants selected in LB with rHAT method

References

1. Metzgar, D. *et al. Acinetobacter* sp. ADP1: an ideal model organism for genetic analysis and genome engineering. *Nucleic Acids Res.* **32**, 5780–90 (2004).
2. Elliott, K. T. & Neidle, E. L. *Acinetobacter baylyi* ADP1: Transforming the choice of model organism. *IUBMB Life* **63**, 1075–1080 (2011).
3. Palmen, R., Vosman, B., Buijsman, P., Breek, C. K. D. & Hellingwerf, K. J. Physiological characterization of natural transformation in *Acinetobacter calcoaceticus*. *J. Gen. Microbiol.* **139**, 295–305 (1993).
4. Palmen, R. & Hellingwerf, K. J. Uptake and processing of DNA by *Acinetobacter calcoaceticus*--a review. *Gene* **192**, 179–90 (1997).
5. Nielsen, K. M., van Elsas, J. D. & Smalla, K. Transformation of *Acinetobacter* sp. strain BD413(pFG4DeltanptII) with transgenic plant DNA in soil microcosms and effects of kanamycin on selection of transformants. *Appl. Environ. Microbiol.* **66**, 1237–42 (2000).
6. Simpson, D. J., Dawson, L. F., Fry, J. C., Rogers, H. J. & Day, M. J. Influence of flanking homology and insert size on the transformation frequency of *Acinetobacter baylyi* BD413. *Environ. Biosafety Res.* **6**, 55–69 (2007).
7. Dubnau, D. DNA Uptake in Bacteria. *Annu. Rev. Microbiol.* **53**, 217–244 (1999).
8. Overballe-Petersen, S. *et al.* Bacterial natural transformation by highly fragmented and damaged DNA. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 19860–19865 (2013).
9. Harms, K., de Vries, J. & Wackernagel, W. A double kill gene cassette for the positive selection of transforming non-selective DNA segments in *Acinetobacter baylyi* BD413. *J. Microbiol. Methods* **69**, 107–115 (2007).
10. Metzgar, D. *et al. Acinetobacter* sp. ADP1: an ideal model organism for genetic analysis and genome engineering. *Nucleic Acids Res.* **32**, 5780–5790 (2004).
11. de Berardinis, V. *et al.* A complete collection of single-gene deletion mutants of *Acinetobacter baylyi* ADP1. *Mol. Syst. Biol.* **4**, 174 (2008).
12. Durot, M. *et al.* Iterative reconstruction of a global metabolic model of *Acinetobacter baylyi* ADP1 using high-throughput growth phenotype and gene essentiality data. *BMC Syst. Biol.* **2**, 85 (2008).
13. Stuani, L. *et al.* Novel metabolic features in *Acinetobacter baylyi* ADP1 revealed by a multiomics approach. *Metabolomics* 1223–1238 (2014). doi:10.1007/s11306-014-0662-x
14. de Berardinis, V., Durot, M., Weissenbach, J. & Salanoubat, M. *Acinetobacter baylyi* ADP1 as a model for metabolic system biology. *Curr. Opin. Microbiol.* **12**, 568–576 (2009).
15. Young, D. M., Parke, D. & Ornston, L. N. Opportunities for genetic investigation afforded by *Acinetobacter baylyi*, a nutritionally versatile bacterial species that is highly competent for natural transformation. *Annu. Rev. Microbiol.* **59**, 519–551 (2005).
16. Leong, C. G. *et al.* The role of core and accessory type IV pilus genes in natural transformation and twitching motility in the bacterium *Acinetobacter baylyi*. *PLoS One* **12**, e0182139 (2017).

17. de Vries, J. & Wackernagel, W. Integration of foreign DNA during natural transformation of *Acinetobacter* sp. by homology-facilitated illegitimate recombination. *Proc. Natl. Acad. Sci.* **99**, 2094–2099 (2002).
18. Hendrickx, L., Hausner, M. & Wuertz, S. Natural genetic transformation in monoculture *Acinetobacter* sp. strain BD413 biofilms. *Appl. Environ. Microbiol.* **69**, 1721–7 (2003).
19. Renda, B. A., Dasgupta, A., Leon, D. & Barrick, J. E. Genome instability mediates the loss of key traits by *Acinetobacter baylyi* ADP1 during laboratory evolution. *J. Bacteriol.* **197**, 872–81 (2015).
20. Hülter, N. *et al.* Costs and benefits of natural transformation in *Acinetobacter baylyi*. *BMC Microbiol.* **17**, 34 (2017).
21. Averhoff, B. & Graf, I. The natural transformation system of *Acinetobacter baylyi* ADP1: A unique DNA transport machinery. in *Acinetobacter: Molecular Biology* (ed. Gerischer, U.) 119–139 (Caister AP, 2008).
22. Harms, K. *et al.* Substitutions of short heterologous DNA segments of intragenomic or extragenomic origins produce clustered genomic polymorphisms. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 15066–15071 (2016).
23. Barbe, V. *et al.* Unique features revealed by the genome sequence of *Acinetobacter* sp. ADP1, a versatile and naturally transformation competent bacterium. *Nucleic Acids Res.* **32**, 5766–79 (2004).
24. Lehtinen, T., Santala, V. & Santala, S. Twin-layer biosensor for real-time monitoring of alkane metabolism. *FEMS Microbiol. Lett.* **364**, (2017).
25. Luo, J., Lehtinen, T., Efimova, E., Santala, V. & Santala, S. Synthetic metabolic pathway for the production of 1-alkenes from lignin-derived molecules. *Microb. Cell Fact.* **18**, 48 (2019).
26. Santala, S. *et al.* Improved triacylglycerol production in *Acinetobacter baylyi* ADP1 by metabolic engineering. *Microb. Cell Fact.* **10**, 36 (2011).
27. Uthoff, S. *et al.* Thio wax ester biosynthesis utilizing the unspecific bifunctional wax ester synthase/acyl coenzyme A:diacylglycerol acyltransferase of *Acinetobacter* sp. strain ADP1. *Appl. Environ. Microbiol.* **71**, 790–6 (2005).
28. Santala, S., Efimova, E., Koskinen, P., Karp, M. T. & Santala, V. Rewiring the wax ester production pathway of *Acinetobacter baylyi* ADP1. *ACS Synth. Biol.* **3**, 145–51 (2014).
29. Kannisto, M. S. *et al.* Metabolic engineering of *Acinetobacter baylyi* ADP1 for removal of *Clostridium butyricum* growth inhibitors produced from lignocellulosic hydrolysates. *Biotechnol. Biofuels* **8**, 198 (2015).
30. Lehtinen, T., Efimova, E., Santala, S. & Santala, V. Improved fatty aldehyde and wax ester production by overexpression of fatty acyl-CoA reductases. *Microb. Cell Fact.* **17**, 1–10 (2018).
31. Huang, W. E. *et al.* Chromosomally located gene fusions constructed in *Acinetobacter* sp. ADP1 for the detection of salicylate. *Environ. Microbiol.* **7**, 1339–48 (2005).
32. DeFraia, C. T., Schmelz, E. A. & Mou, Z. A rapid biosensor-based method for

quantification of free and glucose-conjugated salicylic acid. *Plant Methods* **4**, 28 (2008).
33. Li, C. *et al.* Quantitative measurement of pH influence on SalR regulated gene expression in *Acinetobacter baylyi* ADP1. *J. Microbiol. Methods* **79**, 8–12 (2009).
34. Zhang, D. *et al.* Whole-cell bacterial bioreporter for actively searching and sensing of alkanes and oil spills. *Microb. Biotechnol.* **5**, 87–97 (2012).
35. Song, Y. *et al.* Optimization of Bacterial Whole Cell Bioreporters for Toxicity Assay of Environmental Samples. *Environ. Sci. Technol.* **43**, 7931–7938 (2009).
36. Jiang, B. *et al.* Use of a whole-cell bioreporter, *Acinetobacter baylyi*, to estimate the genotoxicity and bioavailability of chromium(VI)-contaminated soils. *Biotechnol. Lett.* **37**, 343–348 (2015).
37. Ratajczak, A., Geissdörfer, W. & Hillen, W. Expression of alkane hydroxylase from *Acinetobacter* sp. Strain ADP1 is induced by a broad range of n-alkanes and requires the transcriptional activator AlkR. *J. Bacteriol.* **180**, 5822–7 (1998).
38. Kok, R. G., D'Argenio, D. A. & Ornston, L. N. Combining localized PCR mutagenesis and natural transformation in direct genetic analysis of a transcriptional regulator gene, pobR. *J. Bacteriol.* **179**, 4270–6 (1997).
39. Jones, R. M. & Williams, P. A. Mutational analysis of the critical bases involved in activation of the AreR-regulated sigma54-dependent promoter in *Acinetobacter* sp. strain ADP1. *Appl. Environ. Microbiol.* **69**, 5627–35 (2003).
40. Tumen-Velasquez, M. *et al.* Accelerating pathway evolution by increasing the gene dosage of chromosomal segments. *Proc. Natl. Acad. Sci.* **115**, 7105–7110 (2018).
41. Picataggio, S. Potential impact of synthetic biology on the development of microbial systems for the production of renewable fuels and chemicals. *Curr. Opin. Biotechnol.* **20**, 325–9 (2009).
42. Collins and Khalil. Synthetic biology_applications come of age_Collins and Khalil 2010.PDF. (2010).
43. Carothers, J. M., Goler, J. a & Keasling, J. D. Chemical synthesis using synthetic biology. *Curr. Opin. Biotechnol.* **20**, 498–503 (2009).
44. Sleight, S. C., Bartley, B. A., Lieviant, J. A. & Sauro, H. M. Designing and engineering evolutionary robust genetic circuits. *J. Biol. Eng.* **4**, 12 (2010).
45. Renda, B. A., Hammerling, M. J. & Barrick, J. E. Engineering reduced evolutionary potential for synthetic biology. *Mol. Biosyst.* **10**, 1668–1678 (2014).
46. Kelwick, R., MacDonald, J. T., Webb, A. J. & Freemont, P. Developments in the Tools and Methodologies of Synthetic Biology. *Front. Bioeng. Biotechnol.* **2**, 60 (2014).
47. Bölker, M. Complexity in Synthetic Biology: Unnecessary or Essential? in 59–69 (Springer, Cham, 2015). doi:10.1007/978-3-319-02783-8_3
48. McCutcheon, J. P. & Moran, N. a. Extreme genome reduction in symbiotic bacteria. *Nat. Rev. Microbiol.* **10**, 13–26 (2012).
49. Powell, K. How biologists are creating life-like cells from scratch. *Nature* **563**, 172–175 (2018).

50. Chi, H. *et al.* Engineering and modification of microbial chassis for systems and synthetic biology. *Synth. Syst. Biotechnol.* **4**, 25–33 (2019).
51. Calero, P. & Nikel, P. I. Chasing bacterial chassis for metabolic engineering: a perspective review from classical to non-traditional microorganisms. *Microb. Biotechnol.* **12**, 98–124 (2019).
52. Iwadate, Y., Honda, H., Sato, H., Hashimoto, M. & Kato, J. Oxidative stress sensitivity of engineered *Escherichia coli* cells with a reduced genome. *FEMS Microbiol. Lett.* **322**, 25–33 (2011).
53. Reuß, D. R. *et al.* Large-scale reduction of the Bacillus subtilis genome : Consequences for the transcriptional network , resource allocation , and metabolism. 1–31 (2017). doi:10.1101/gr.215293.116
54. Hutchison, C. A. *et al.* Design and synthesis of a minimal bacterial genome. *Science* **351**, aad6253 (2016).
55. Pósfai, G. *et al.* Emergent properties of reduced-genome *Escherichia coli*. *Science (80-.).* **312**, 1044–1046 (2006).
56. Park, M. K. *et al.* Enhancing recombinant protein production with an *Escherichia coli* host strain lacking insertion sequences. *Appl. Microbiol. Biotechnol.* **98**, 6701–6713 (2014).
57. Westers, H. *et al.* Genome Engineering Reveals Large Dispensable Regions in Bacillus subtilis. *Mol. Biol. Evol.* **20**, 2076–2090 (2003).
58. Umenhoffer, K. *et al.* Reduced evolvability of *Escherichia coli* MDS42, an IS-less cellular chassis for molecular and synthetic biology applications. *Microb. Cell Fact.* **9**, 38 (2010).
59. Suárez, G. A., Renda, B. A., Dasgupta, A. & Barrick, J. E. Reduced mutation rate and increased transformability of transposon-free *Acinetobacter baylyi* ADP1-ISx. *Appl. Environ. Microbiol.* **83**, e01025-17 (2017).
60. Baumgart, M. *et al.* Construction of a prophage-free variant of *Corynebacterium glutamicum* ATCC 13032 for use as a platform strain for basic research and industrial biotechnology. *Appl. Environ. Microbiol.* **79**, 6006–15 (2013).
61. Aguilar Suárez, R., Stülke, J. & van Dijl, J. M. Less Is More: Toward a Genome-Reduced *Bacillus* Cell Factory for "Difficult Proteins". *ACS Synth. Biol.* **8**, 99–108 (2019).
62. Li, Y. *et al.* Characterization of genome-reduced *Bacillus subtilis* strains and their application for the production of guanosine and thymidine. *Microb. Cell Fact.* **15**, 94 (2016).
63. Martínez-García, E., Nikel, P. I., Aparicio, T. & de Lorenzo, V. *Pseudomonas* 2.0: genetic upgrading of *P. putida* KT2440 as an enhanced host for heterologous gene expression. *Microb. Cell Fact.* **13**, 159 (2014).
64. Lieder, S., Nikel, P. I., de Lorenzo, V. & Takors, R. Genome reduction boosts heterologous gene expression in *Pseudomonas putida*. *Microb. Cell Fact.* **14**, 23 (2015).
65. Mizoguchi, H., Sawano, Y., Kato, J. & Mori, H. Superpositioning of deletions promotes growth of *Escherichia coli* with a reduced genome. *DNA Res.* **15**, 277–

84 (2008).
66. Lee, J. H. *et al.* Metabolic engineering of a reduced-genome strain of *Escherichia coli* for L-threonine production. *Microb. Cell Fact.* **8**, 2 (2009).
67. Komatsu, M., Uchiyama, T., Omura, S., Cane, D. E. & Ikeda, H. Genome-minimized *Streptomyces* host for the heterologous expression of secondary metabolism. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 2646–51 (2010).
68. Mizoguchi, H., Sawano, Y., Kato, J. I. & Mori, H. Superpositioning of deletions promotes growth of *Escherichia coli* with a reduced genome. *DNA Res.* **15**, 277–284 (2008).
69. Morimoto, T. *et al.* Enhanced recombinant protein productivity by genome reduction in *Bacillus subtilis*. *DNA Res.* **15**, 73–81 (2008).
70. Ara, K. *et al.* Bacillus minimum genome factory: effective utilization of microbial genome information. *Biotechnol. Appl. Biochem.* **46**, 169 (2007).
71. Karcagi, I. *et al.* Indispensability of Horizontally Transferred Genes and Its Impact on Bacterial Genome Streamlining. *Mol. Biol. Evol.* **33**, 1257–1269 (2016).
72. Renda, B. A., Hammerling, M. J. & Barrick, J. E. Engineering reduced evolutionary potential for synthetic biology. *Mol. Biosyst.* **10**, 1668–78 (2014).
73. Posfai, G. *et al.* Emergent Properties of Reduced-Genome *Escherichia coli*. *Science (80-.).* **312**, 1044–1046 (2006).
74. Manabe, K. *et al.* Combined Effect of Improved Cell Yield and Increased Specific Productivity Enhances Recombinant Enzyme Production in Genome-Reduced Bacillus subtilis Strain MGB874. *Appl. Environ. Microbiol.* **77**, 8370–8381 (2011).
75. Rahmen, N. *et al.* Exchange of single amino acids at different positions of a recombinant protein affects metabolic burden in *Escherichia coli*. *Microb. Cell Fact.* **14**, 10 (2015).
76. Borkowski, O., Ceroni, F., Stan, G.-B. & Ellis, T. Overloaded and stressed: whole-cell considerations for bacterial synthetic biology. *Curr. Opin. Microbiol.* **33**, 123–130 (2016).
77. Wu, G. *et al.* Metabolic Burden: Cornerstones in Synthetic Biology and Metabolic Engineering Applications. *Trends Biotechnol.* **34**, 652–664 (2016).
78. Lieder, S., Nikel, P. I., de Lorenzo, V. & Takors, R. Genome reduction boosts heterologous gene expression in *Pseudomonas putida*. *Microb. Cell Fact.* **14**, 1–14 (2015).
79. Unthan, S. *et al.* Chassis organism from Corynebacterium glutamicum - a top-down approach to identify and delete irrelevant gene clusters. *Biotechnol. J.* **10**, 290–301 (2015).
80. Martínez-García, E. & de Lorenzo, V. The quest for the minimal bacterial genome. *Curr. Opin. Biotechnol.* **42**, 216–224 (2016).
81. Juhas, M., Reuss, D. R., Zhu, B. & Commichau, F. M. Bacillus subtilis and *Escherichia coli* essential genes and minimal cell factories after one decade of genome engineering. *Microbiology* **160**, 2341–2351 (2014).
82. Kurokawa, M., Seno, S., Matsuda, H. & Ying, B.-W. Correlation between genome

reduction and bacterial growth. *DNA Res.* **23**, 517–525 (2016).
83. Hashimoto, M. *et al.* Cell size and nucleoid organization of engineered *Escherichia coli* cells with a reduced genome. *Mol. Microbiol.* **55**, 137–149 (2004).
84. Hutchison, C. A. *et al.* Design and synthesis of a minimal bacterial genome. *Science* **351**, aad6253 (2016).
85. Hirokawa, Y. *et al.* Genetic manipulations restored the growth fitness of reduced-genome *Escherichia coli*. *J. Biosci. Bioeng.* **116**, 52–58 (2013).
86. Choe, D. *et al.* Adaptive laboratory evolution of a genome-reduced *Escherichia coli*. *Nat. Commun.* **10**, (2019).
87. Rancati, G., Moffat, J., Typas, A. & Pavelka, N. Emerging and evolving concepts in gene essentiality. *Nat. Rev. Genet.* **19**, 34–49 (2018).
88. Wang, L. & Maranas, C. D. MinGenome: An *In Silico* Top-Down Approach for the Synthesis of Minimized Genomes. *ACS Synth. Biol.* **7**, 462–473 (2018).
89. Morimoto, T. *et al.* Enhanced Recombinant Protein Productivity by Genome Reduction in Bacillus subtilis. *DNA Res.* **15**, 73–81 (2008).
90. Manabe, K. *et al.* Combined Effect of Improved Cell Yield and Increased Specific Productivity Enhances Recombinant Enzyme Production in Genome-Reduced Bacillus subtilis Strain MGB874. *Appl. Environ. Microbiol.* **77**, 8370–8381 (2011).
91. Reuß, D. R. *et al.* Large-scale reduction of the Bacillus subtilis genome: consequences for the transcriptional network, resource allocation, and metabolism. *Genome Res.* **27**, 289–299 (2017).
92. Metzgar, D. *et al. Acinetobacter* sp. ADP1: an ideal model organism for genetic analysis and genome engineering. *Nucleic Acids Res.* **32**, 5780–90 (2004).
93. Vaneechoutte, M. *et al.* Naturally transformable *Acinetobacter* sp. strain ADP1 belongs to the newly described species *Acinetobacter baylyi*. *Appl. Environ. Microbiol.* **72**, 932–936 (2006).
94. Elliott, K. T. & Neidle, E. L. *Acinetobacter baylyi* ADP1: transforming the choice of model organism. *IUBMB Life* **63**, 1075–1080 (2011).
95. de Berardinis, V., Durot, M., Weissenbach, J. & Salanoubat, M. *Acinetobacter baylyi* ADP1 as a model for metabolic system biology. *Curr. Opin. Microbiol.* **12**, 568–576 (2009).
96. Young, D. M., Parke, D. & Ornston, L. N. Opportunities for genetic investigation afforded by *Acinetobacter baylyi*, a nutritionally versatile bacterial species that is highly competent for natural transformation. *Annu. Rev. Microbiol.* **59**, 519–551 (2005).
97. Zhang, D. *et al.* Whole-cell bacterial bioreporter for actively searching and sensing of alkanes and oil spills. *Microb. Biotechnol.* **5**, 87–97 (2012).
98. Buchan, A. & Ornston, L. N. When coupled to natural transformation in *Acinetobacter* sp. strain ADP1, PCR mutagenesis is made less random by mismatch repair. *Appl. Environ. Microbiol.* **71**, 7610–2 (2005).
99. Vandecraen, J., Chandler, M., Aertsen, A. & Van Houdt, R. The impact of

insertion sequences on bacterial genome plasticity and adaptability. *Crit. Rev. Microbiol.* **0**, 1–22 (2017).
100. Raeside, C. *et al.* Large chromosomal rearrangements during a long-term evolution experiment with *Escherichia coli*. *MBio* **5**, (2014).
101. Lee, H., Doak, T. G., Popodi, E., Foster, P. L. & Tang, H. Insertion sequence-caused large-scale rearrangements in the genome of *Escherichia coli*. *Nucleic Acids Res.* **44**, 7109–19 (2016).
102. Barbe, V. *et al.* Unique features revealed by the genome sequence of *Acinetobacter* sp. ADP1, a versatile and naturally transformation competent bacterium. *Nucleic Acids Res.* **32**, 5766–79 (2004).
103. Gerischer, U., D'Argenio, D. A. & Ornston, L. N. IS*1236*, a newly discovered member of the IS*3* family, exhibits varied patterns of insertion into the *Acinetobacter calcoaceticus* chromosome. *Microbiology* **142**, 1825–1831 (1996).
104. Mahillon, J. & Chandler, M. Insertion sequences. *Microbiol. Mol. Biol. Rev.* **62**, 725–774 (1998).
105. Segura, A., Bünz, P. V, D'Argenio, D. A. & Ornston, L. N. Genetic analysis of a chromosomal region containing *vanA* and *vanB*, genes required for conversion of either ferulate or vanillate to protocatechuate in *Acinetobacter*. *J. Bacteriol.* **181**, 3494–3504 (1999).
106. Cuff, L. E. *et al.* Analysis of IS*1236*-mediated gene amplification events in *Acinetobacter baylyi* ADP1. *J. Bacteriol.* **194**, 4395–405 (2012).
107. Jezequel, N., Lagomarsino, M. C., Heslot, F. & Thomen, P. Long-term diversity and genome adaptation of *Acinetobacter baylyi* in a minimal-medium chemostat. *Genome Biol. Evol.* **5**, 87–97 (2013).
108. Renda, B. A., Dasgupta, A., Leon, D. & Barrick, J. E. Genome instability mediates the loss of key traits by *Acinetobacter baylyi* ADP1 during laboratory evolution. *J. Bacteriol.* **197**, 872–81 (2015).
109. Bacher, J. M., Metzgar, D. & de Crécy-Lagard, V. Rapid evolution of diminished transformability in *Acinetobacter baylyi*. *J. Bacteriol.* **188**, 8534–42 (2006).
110. Seaton, S. C. *et al.* Genome-wide selection for increased copy number in *Acinetobacter baylyi* ADP1: locus and context-dependent variation in gene amplification. *Mol. Microbiol.* **83**, 520–535 (2012).
111. Deatherage, D. E. & Barrick, J. E. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using *breseq*. *Methods Mol. Biol.* **1151**, 165–188 (2014).
112. Barrick, J. E. *et al.* Identifying structural variation in haploid microbial genomes from short-read resequencing data using *breseq*. *BMC Genomics* **15**, 1039 (2014).
113. R Core Team. *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, 2016).
114. de Berardinis, V. *et al.* A complete collection of single-gene deletion mutants of *Acinetobacter baylyi* ADP1. *Mol. Syst. Biol.* **4**, 174 (2008).
115. Hare, J. M., Perkins, S. N. & Gregg-Jolly, L. A. A constitutively expressed, truncated *umuDC* operon regulates the *recA*-dependent DNA damage induction of

a gene in *Acinetobacter baylyi* strain ADP1. *Appl. Environ. Microbiol.* **72**, 4036–43 (2006).
116. Hall, B. M., Ma, C.-X., Liang, P. & Singh, K. K. Fluctuation analysis CalculatOR: a web tool for the determination of mutation rate using Luria-Delbrück fluctuation analysis. *Bioinformatics* **25**, 1564–1565 (2009).
117. Murin, C. D., Segal, K., Bryksin, A. & Matsumura, I. Expression vectors for *Acinetobacter baylyi* ADP1. *Appl. Environ. Microbiol.* **78**, 280–283 (2012).
118. Zheng, Q. Comparing mutation rates under the Luria-Delbrück protocol. *Genetica* **144**, 351–9 (2016).
119. Zheng, Q. rSalvador 1.7: an R tool for the Luria-Delbrück fluctuation assay. (2017). Available at: http://eeeeeric.github.io/rSalvador.
120. Barrick, J. E. & Lenski, R. E. Genome dynamics during experimental evolution. *Nat. Rev. Genet.* **14**, 827–839 (2013).
121. Lee, H., Popodi, E., Tang, H. & Foster, P. L. Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proc. Natl. Acad. Sci. U. S. A.* **109**, E2774–E2783 (2012).
122. Dillon, M. M.; Sung, W.; Lynch, M. . C. V. S. Genome-wide biases in the rate and molecular spectrum of spontaneous mutations in *Vibrio cholerae* and *Vibrio fischeri*. *Rev.* **34**, 93–109 (2016).
123. Dettman, J. R., Sztepanacz, J. L. & Kassen, R. The properties of spontaneous mutations in the opportunistic pathogen *Pseudomonas aeruginosa*. *BMC Genomics* **17**, 1–14 (2016).
124. Wösten, M. M. S. M. Eubacterial sigma-factors. *FEMS Microbiol. Rev.* **22**, (1998).
125. Bekker, M., De Vries, S., Ter Beek, A., Hellingwerf, K. J. & Teixeira De Mattos, M. J. Respiration of *Escherichia coli* can be fully uncoupled via the nonelectrogenic terminal cytochrome *bd*-II oxidase. *J. Bacteriol.* **191**, 5510–5517 (2009).
126. Lázár, V. *et al.* Bacterial evolution of antibiotic hypersensitivity. *Mol. Syst. Biol.* **9**, 700 (2013).
127. Jack, B. R. *et al.* Predicting the genetic stability of engineered DNA sequences with the EFM Calculator. *ACS Synth. Biol.* **4**, 939–943 (2015).
128. Kickstein, E., Harms, K. & Wackernagel, W. Deletions of *recBCD* or *recD* influence genetic transformation differently and are lethal together with a *recJ* deletion in *Acinetobacter baylyi*. *Microbiology* **153**, 2259–70 (2007).
129. Csörgo, B., Fehér, T., Tímár, E., Blattner, F. R. & Pósfai, G. Low-mutation-rate, reduced-genome *Escherichia coli*: an improved host for faithful maintenance of engineered genetic constructs. *Microb. Cell Fact.* **11**, 11 (2012).
130. Hare, J. M., Adhikari, S., Lambert, K. V, Hare, A. E. & Grice, A. N. The *Acinetobacter* regulatory UmuDAb protein cleaves in response to DNA damage with chimeric LexA/UmuD characteristics. *FEMS Microbiol. Lett.* **334**, 57–65 (2012).
131. Hare, J. M., Ferrell, J. C., Witkowski, T. A. & Grice, A. N. Prophage induction and differential RecA and UmuDAb transcriptome regulation in the DNA damage

responses of *Acinetobacter baumannii* and *Acinetobacter baylyi*. *PLoS One* **9**, e93861 (2014).
132. Newmark, K. G., O'Reilly, E. K., Pohlhaus, J. R. & Kreuzer, K. N. Genetic analysis of the requirements for SOS induction by nalidixic acid in *Escherichia coli*. *Gene* **356**, 69–76 (2005).
133. Pósfai, G. *et al.* Emergent properties of reduced-genome *Escherichia coli*. *Science* **312**, 1044–6 (2006).
134. Choi, J. W., Yim, S. S., Kim, M. J. & Jeong, K. J. Enhanced production of recombinant proteins with *Corynebacterium glutamicum* by deletion of insertion sequences (IS elements). *Microb. Cell Fact.* **14**, 207 (2015).
135. Zhu, D. *et al.* Enhanced heterologous protein productivity by genome reduction in *Lactococcus lactis* NZ9000. *Microb. Cell Fact.* **16**, 1 (2017).
136. Reuß, D. R. *et al.* Large-scale reduction of the *Bacillus subtilis* genome: consequences for the transcriptional network, resource allocation, and metabolism. *Genome Res.* **27**, 289–299 (2017).
137. Rahmer, R., Heravi, K. M. & Altenbuchner, J. Construction of a super-competent *Bacillus subtilis* 168 using the PmtlA-comKS inducible cassette. *Front. Microbiol.* **6**, 1–11 (2015).
138. Arkin, A. P. & Fletcher, D. A. Fast, cheap and somewhat in control. *Genome Biol.* **7**, 114 (2006).
139. Juhas, M., Reuß, D. R., Zhu, B. & Commichau, F. M. Bacillus subtilis and *Escherichia coli* essential genes and minimal cell factories after one decade of genome engineering. *Microbiol. (United Kingdom)* **160**, 2341–2351 (2014).
140. Calero, P. & Nikel, P. I. Chasing bacterial chassis for metabolic engineering: a perspective review from classical to non-traditional microorganisms. *Microb. Biotechnol.* **12**, 98–124 (2019).
141. Fehér, T., Papp, B., Pal, C. & Pósfai, G. Systematic genome reductions: theoretical and experimental approaches. *Chem. Rev.* **107**, 3498–3513 (2007).
142. Leprince, A., van Passel, M. W. J. & dos Santos, V. a P. M. Streamlining genomes: toward the generation of simplified and stabilized microbial systems. *Curr. Opin. Biotechnol.* **23**, 651–8 (2012).
143. Glass, J. I., Merryman, C., Wise, K. S., Iii, C. A. H. & Smith, H. O. Minimal Cells — Real and Imagined. 1–12 (2017).
144. Martínez-García, E. & de Lorenzo, V. The quest for the minimal bacterial genome. *Curr. Opin. Biotechnol.* **42**, 216–224 (2016).
145. Posfai, G. *et al.* Emergent Properties of Reduced-Genome *Escherichia coli*. *Science (80-.).* **312**, 1044–1046 (2006).
146. Suárez, G. A., Renda, B. A., Dasgupta, A. & Barrick, J. E. Reduced Mutation Rate and Increased Transformability of Transposon-Free *Acinetobacter baylyi* ADP1-ISx. *Appl. Environ. Microbiol.* **83**, (2017).
147. Choi, J. W., Yim, S. S., Kim, M. J. & Jeong, K. J. Enhanced production of recombinant proteins with *Corynebacterium glutamicum* by deletion of insertion sequences (IS elements). *Microb. Cell Fact.* **14**, 207 (2015).

148. Martínez-García, E., Nikel, P. I., Chavarría, M. & de Lorenzo, V. The metabolic cost of flagellar motion in *Pseudomonas putida* KT2440. *Environ. Microbiol.* **16**, 291–303 (2014).
149. Wang, Z. *et al.* Developing genome-reduced *Pseudomonas chlororaphis* strains for the production of secondary metabolites. *BMC Genomics* **18**, 1–14 (2017).
150. Rancati, G., Moffat, J., Typas, A. & Pavelka, N. Emerging and evolving concepts in gene essentiality. *Nat. Rev. Genet.* **19**, 34–49 (2018).
151. Seaton, S. C. *et al.* Genome-wide selection for increased copy number in *Acinetobacter baylyi* ADP1: locus and context-dependent variation in gene amplification. *Mol. Microbiol.* **83**, 520–535 (2012).
152. Renda, B. A., Chan, C., Parent, K. N. & Barrick, J. E. Emergence of a Competence-Reducing Filamentous Phage from the Genome of *Acinetobacter baylyi* ADP1. *J. Bacteriol.* **198**, 3209–3219 (2016).
153. Ornston, N., Parke, D. & Ornston, L. N. Opportunities for genetic investigation afforded by *Acinetobacter baylyi*, a nutritionally versatile bacterial species that is highly competent for natural transformation. *Annu. Rev. Microbiol.* **59**, annurev.mi.59.082405.100001 (2005).
154. Fischer, R., Bleichrodt, F. S. & Gerischer, U. C. Aromatic degradative pathways in *Acinetobacter baylyi* underlie carbon catabolite repression. *Microbiology* **154**, 3095–3103 (2008).
155. Barbe, V. *et al.* Unique features revealed by the genome sequence of *Acinetobacter* sp. ADP1, a versatile and naturally transformation competent bacterium. *Nucleic Acids Res.* **32**, 5766–5779 (2004).
156. de Berardinis, V., Durot, M., Weissenbach, J. & Salanoubat, M. *Acinetobacter baylyi* ADP1 as a model for metabolic system biology. *Curr. Opin. Microbiol.* **12**, 568–576 (2009).
157. Stuani, L. *et al.* Novel metabolic features in *Acinetobacter baylyi* ADP1 revealed by a multiomics approach. *Metabolomics* **10**, 1223–1238 (2014).
158. Santala, S. *et al.* Improved Triacylglycerol Production in *Acinetobacter baylyi* ADP1 by Metabolic Engineering. *Microb. Cell Fact.* **10**, 36 (2011).
159. Jones, R. M. & Williams, P. A. Mutational analysis of the critical bases involved in activation of the AreR-regulated sigma54-dependent promoter in *Acinetobacter* sp. strain ADP1. *Appl. Environ. Microbiol.* **69**, 5627–35 (2003).
160. Tumen-Velasquez, M. *et al.* Accelerating pathway evolution by increasing the gene dosage of chromosomal segments. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 7105–7110 (2018).
161. Engler, C., Kandzia, R. & Marillonnet, S. A One Pot, One Step, Precision Cloning Method with High Throughput Capability. *PLoS One* **3**, e3647 (2008).
162. Lee, M. E., DeLoache, W. C., Cervantes, B. & Dueber, J. E. A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. *ACS Synth. Biol.* **4**, 975–986 (2015).
163. Engler, C., Kandzia, R. & Marillonnet, S. A one pot, one step, precision cloning method with high throughput capability. *PLoS One* **3**, (2008).

164. Lee, M. E., DeLoache, W. C., Cervantes, B. & Dueber, J. E. A highly characterized yeast toolkit for modular, multipart assembly. *ACS Synth. Biol.* **4**, 975–986 (2015).
165. R Core Team. *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, 2018).
166. Kulasekara, H. D. *et al.* A novel two-component system controls the expression of *Pseudomonas aeruginosa* fimbrial cup genes. *Mol. Microbiol.* **55**, 368–380 (2004).
167. Demarre, G. *et al.* A new family of mobilizable suicide plasmids based on broad host range R388 plasmid (IncW) and RP4 plasmid (IncPα) conjugative machineries and their cognate *Escherichia coli* host strains. *Res. Microbiol.* **156**, 245–255 (2005).
168. Powell, J. E., Leonard, S. P., Kwong, W. K., Engel, P. & Moran, N. A. Genome-wide screen identifies host colonization determinants in a bacterial gut symbiont. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 13887–13892 (2016).
169. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
170. Huerta-Cepas, J. *et al.* Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol. Biol. Evol.* **34**, 2115–2122 (2017).
171. Renda, B. A., Dasgupta, A., Leon, D. & Barrick, J. E. Genome instability mediates the loss of key traits by *Acinetobacter baylyi* ADP1 during laboratory evolution. *J. Bacteriol.* **197**, 872–81 (2015).
172. Barrick, J. E. *et al.* Identifying structural variation in haploid microbial genomes from short-read resequencing data using breseq. *BMC Genomics* **15**, 1039 (2014).
173. Deatherage, D. E. & Barrick, J. E. Identification of Mutations in Laboratory-Evolved Microbes from Next-Generation Sequencing Data Using breseq. in *Methods in molecular biology (Clifton, N.J.)* **1151**, 165–188 (2014).
174. Overballe-Petersen, S. *et al.* Bacterial natural transformation by highly fragmented and damaged DNA. *Proc. Natl. Acad. Sci.* **110**, 19860–19865 (2013).
175. Simpson, D. J., Dawson, L. F., Fry, J. C., Rogers, H. J. & Day, M. J. Influence of flanking homology and insert size on the transformation frequency of *Acinetobacter baylyi* BD413. *Environ. Biosafety Res.* **6**, 55–69 (2007).
176. Leonard, S. P. *et al.* Genetic engineering of bee gut microbiome bacteria with a toolkit for modular assembly of broad-host-range plasmids. *ACS Synth. Biol.* **7**, 1279–1290 (2018).
177. de Berardinis, V. *et al.* A complete collection of single-gene deletion mutants of *Acinetobacter baylyi* ADP1. *Mol. Syst. Biol.* **4**, 174 (2008).
178. Baba, T. *et al.* Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2**, 2006.0008 (2006).
179. Yamamoto, N. *et al.* Update on the Keio collection of *Escherichia coli* single-gene deletion mutants. *Mol. Syst. Biol.* **5**, 335 (2009).
180. Goodall, E. C. A. *et al.* The Essential Genome of *Escherichia coli* K-12. (2018). doi:10.1128/mBio.02096
181. Reams, A. B. & Neidle, E. L. Gene amplification involves site-specific short

homology-independent illegitimate recombination in *Acinetobacter* sp. strain ADP1. *J. Mol. Biol.* **338**, 643–56 (2004).
182. Seaton, S. C. *et al.* Genome-wide selection for increased copy number in *Acinetobacter baylyi* ADP1: locus and context-dependent variation in gene amplification. *Mol. Microbiol.* **83**, 520–535 (2012).
183. de Berardinis, V. *et al.* A complete collection of single-gene deletion mutants of *Acinetobacter baylyi* ADP1. *Mol. Syst. Biol.* **4**, 174 (2008).
184. Touchon, M. *et al.* The genomic diversification of the whole *Acinetobacter* genus: origins, mechanisms, and consequences. *Genome Biol. Evol.* **6**, 2866–2882 (2014).
185. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**, 733–40 (2009).
186. Karah, N. *et al.* CRISPR-cas subtype I-Fb in *Acinetobacter baumannii*: evolution and utilization for strain subtyping. *PLoS One* **10**, e0118205 (2015).
187. Biswas, A., Gagnon, J. N., Brouns, S. J. J., Fineran, P. C. & Brown, C. M. CRISPRTarget: Bioinformatic prediction and analysis of crRNA targets. *RNA Biol.* **10**, 817–827 (2013).
188. Mori, H. *et al.* Identification of Essential Genes and Synthetic Lethal Gene Combinations in *Escherichia coli* K-12. in 45–65 (Humana Press, New York, NY, 2015). doi:10.1007/978-1-4939-2398-4_4
189. Typas, A. *et al.* High-throughput, quantitative analyses of genetic interactions in E. coli. *Nat. Methods* **5**, 781–7 (2008).
190. Freed, E. *et al.* Building a genome engineering toolbox in nonmodel prokaryotic microbes. *Biotechnol. Bioeng.* **115**, 2120–2138 (2018).
191. Sharma, R. *et al.* Identification of novel regulatory small RNAs in *Acinetobacter baumannii*. *PLoS One* **9**, (2014).
192. Wang, L. & Maranas, C. D. MinGenome: An *in silico* top-down approach for the synthesis of minimized genomes. *ACS Synth. Biol.* **7**, 462–473 (2018).
193. Makarova, K. S. *et al.* Evolution and classification of the CRISPR-Cas systems. *Nat. Rev. Microbiol.* **9**, 467–77 (2011).
194. Geng, P., Leonard, S. P., Mishler, D. M. & Barrick, J. E. Synthetic Genome Defenses against Selfish DNA Elements Stabilize Engineered Bacteria against Evolutionary Failure. *ACS Synth. Biol.* **1**, 1–29 (2019).
195. Luo, M. L., Mullis, A. S., Leenay, R. T. & Beisel, C. L. Repurposing endogenous type I CRISPR-Cas systems for programmable gene repression. *Nucleic Acids Res.* **43**, 674–81 (2015).
196. Eisenstein, M. Pursuing the simple life. *Nat. Methods* **14**, 117–121 (2017).
197. Akeno, Y., Ying, B.-W., Tsuru, S. & Yomo, T. A reduced genome decreases the host carrying capacity for foreign DNA. *Microb. Cell Fact.* **13**, 49 (2014).
198. Sharma, S. S., Campbell, J. W., Frisch, D., Blattner, F. R. & Harcum, S. W. Expression of two recombinant chloramphenicol acetyltransferase variants in highly reduced genome *Escherichia coli* strains. *Biotechnol. Bioeng.* **98**, 1056–1070 (2007).

199. Csorgo, B., Feher, T., Timar, E., Blattner, F. R. & Posfai, G. Low-mutation-rate, reduced-genome *Escherichia coli*: An improved host for faithful maintenance of engineered genetic constructs. *Microb. Cell Fact.* **11**, 11 (2012).
200. Martínez-García, E., Nikel, P. I., Aparicio, T. & de Lorenzo, V. *Pseudomonas* 2.0: genetic upgrading of *P. putida* KT2440 as an enhanced host for heterologous gene expression. *Microb. Cell Fact.* **13**, 159 (2014).
201. Hashimoto, M. *et al.* Cell size and nucleoid organization of engineered *Escherichia coli* cells with a reduced genome. *Mol. Microbiol.* **55**, 137–149 (2005).
202. Nilsson, A. I. *et al.* From The Cover: Bacterial genome size reduction by experimental evolution. *Proc. Natl. Acad. Sci.* **102**, 12112–12116 (2005).
203. Bjorkman, J., Hughes, D. & Andersson, D. I. Virulence of antibiotic-resistant Salmonella typhimurium. *Proc. Natl. Acad. Sci.* **95**, 3949–3953 (1998).
204. Watanabe, K., Tominaga, K., Kitamura, M. & Kato, J. Systematic identification of synthetic lethal mutations with reduced-genome *Escherichia coli*: synthetic genetic interactions among *yoaA*, *xthA* and *holC* related to survival from MMS. *Genes Genet. Syst.* **91**, 183–188 (2016).
205. Yuan, X. *et al.* Single-Cell Microfluidics to Study the Effects of Genome Deletion on Bacterial Growth Behavior. *ACS Synth. Biol.* **6**, 2219–2227 (2017).
206. O'Connor, T. J., Adepoju, Y., Boyd, D. & Isberg, R. R. Minimization of the Legionella pneumophila genome reveals chromosomal regions involved in host range expansion. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 14733–40 (2011).
207. Conrad, T. M. *et al.* RNA polymerase mutants found through adaptive evolution reprogram *Escherichia coli* for optimal growth in minimal media. *Proc. Natl. Acad. Sci.* **107**, 20500–20505 (2010).
208. McCloskey, D. *et al.* Evolution of gene knockout strains of E. coli reveal regulatory architectures governed by metabolism. *Nat. Commun.* **9**, 3796 (2018).
209. Blank, D., Wolf, L., Ackermann, M. & Silander, O. K. The predictability of molecular evolution during functional innovation. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 3044–9 (2014).
210. Bergmiller, T., Ackermann, M. & Silander, O. K. Patterns of Evolutionary Conservation of Essential Genes Correlate with Their Compensability. *PLoS Genet.* **8**, e1002803 (2012).
211. Patrick, W. M., Quandt, E. M., Swartzlander, D. B. & Matsumura, I. Multicopy suppression underpins metabolic evolvability. *Mol. Biol. Evol.* **24**, 2716–22 (2007).
212. Wiser, M. J., Ribeck, N. & Lenski, R. E. Long-Term Dynamics of Adaptation in Asexual Populations. *Science (80-.).* **342**, 1364–1367 (2013).
213. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
214. Barbe, V. *et al.* Unique features revealed by the genome sequence of *Acinetobacter* sp. ADP1, a versatile and naturally transformation competent bacterium. *Nucleic Acids Res.* **32**, 5766–5779 (2004).

215. Deatherage, D. E., Traverse, C. C., Wolf, L. N. & Barrick, J. E. Detecting rare structural variation in evolving microbial populations from new sequence junctions using breseq. *Front. Genet.* **5**, 468 (2014).
216. Krzywinski, M. *et al.* Circos: An information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).
217. Barrick, J. E., Kauth, M. R., Strelioff, C. C. & Lenski, R. E. *Escherichia coli* rpoB Mutants Have Increased Evolvability in Proportion to Their Fitness Defects. *Mol. Biol. Evol.* **27**, 1338–1347 (2010).
218. Rojas Echenique, J. I., Kryazhimskiy, S., Nguyen Ba, A. N. & Desai, M. M. Modular epistasis and the compensatory evolution of gene deletion mutants. *PLOS Genet.* **15**, e1007958 (2019).
219. Harcombe, W., Springman, R. & Bull, J. Compensatory evolution for a gene deletion is not limited to its immediate functional network. *BMC Evol. Biol.* **9**, 106 (2009).
220. Szamecz, B. *et al.* The Genomic Landscape of Compensatory Evolution. *PLoS Biol.* **12**, e1001935 (2014).
221. Teng, X. *et al.* Genome-wide Consequences of Deleting Any Single Gene. *Mol. Cell* **52**, 485–494 (2013).
222. Vaneechoutte, M. *et al.* Oil-degrading *Acinetobacter* strain RAG-1 and strains described as '*Acinetobacter venetianus* sp. nov.' belong to the same genomic species. *Res. Microbiol.* **150**, 69–73 (1999).
223. Gutnick, D. L. & Nakar, D. Analysis of the wee gene cluster responsible for the biosynthesis of the polymeric bioemulsifier from the oil-degrading strain *Acinetobacter lwoffii* RAG-1. *Microbiology* **147**, 1937–1946 (2001).
224. KAPLAN, N., ROSENBERG, E., JANN, B. & JANN, K. Structural studies of the capsular polysaccharide of *Acinetobacter calcoaceticus* BD4. *Eur. J. Biochem.* **152**, 453–458 (1985).
225. Li, Z., Pandit, S. & Deutscher, M. P. 3' exoribonucleolytic trimming is a common feature of the maturation of small, stable RNAs in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 2856–61 (1998).
226. Zuo, Y., Wang, Y. & Malhotra, A. Crystal Structure of *Escherichia coli* RNase D, an Exoribonuclease Involved in Structured RNA Processing. *Structure* **13**, 973–984 (2005).
227. Blouin, R. T., Zaniewski, R. & Deutscher, M. P. Ribonuclease D is not essential for the normal growth of *Escherichia coli* or bacteriophage T4 or for the biosynthesis of a T4 suppressor tRNA. *J. Biol. Chem.* **258**, 1423–6 (1983).
228. Zhang, J. R. & Deutscher, M. P. Transfer RNA is a substrate for RNase D in vivo. *J. Biol. Chem.* **263**, 17909–12 (1988).
229. Kelly, K. O. & Deutscher, M. P. The presence of only one of five exoribonucleases is sufficient to support the growth of *Escherichia coli*. *J. Bacteriol.* **174**, 6682–6684 (1992).
230. Wright, M. S., Jacobs, M. R., Bonomo, R. A. & Adams, M. D. Transcriptome Remodeling of *Acinetobacter baumannii* during Infection and Treatment. *MBio* **8**,

e02193-16 (2017).
231. Kalvari, I. *et al.* Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Res.* **46**, D335–D342 (2018).
232. Mann, M., Wright, P. R. & Backofen, R. IntaRNA 2.0: enhanced and customizable prediction of RNA–RNA interactions. *Nucleic Acids Res.* **45**, W435–W439 (2017).
233. Sabnis, N. A., Yang, H. & Romeo, T. Pleiotropic Regulation of Central Carbohydrate Metabolism in *Escherichia coli* via the Gene *csrA*. *J. Biol. Chem.* **270**, 29096–29104 (1995).
234. Pannuri, A. *et al.* Circuitry Linking the Catabolite Repression and Csr Global Regulatory Systems of *Escherichia coli*. *J. Bacteriol.* **198**, 3000–3015 (2016).
235. Esquerré, T. *et al.* The Csr system regulates genome-wide mRNA stability and transcription and thus gene expression in *Escherichia coli*. *Sci. Rep.* **6**, 25057 (2016).
236. Darmon, E., Eykelenboom, J. K., Lopez-Vernaza, M. A., White, M. A. & Leach, D. R. F. Repair on the Go: E. coli Maintains a High Proliferation Rate while Repairing a Chronic DNA Double-Strand Break. *PLoS One* **9**, e110784 (2014).
237. Deatherage, D. E., Leon, D., Rodriguez, Á. E., Omar, S. K. & Barrick, J. E. Directed evolution of *Escherichia coli* with lower-than-natural plasmid mutation rates. *Nucleic Acids Res.* **46**, 9236–9250 (2018).
238. Renda, B. A., Chan, C., Parent, K. N. & Barrick, J. E. Emergence of a Competence-Reducing Filamentous Phage from the Genome of *Acinetobacter baylyi* ADP1. *J. Bacteriol.* **198**, 3209–3219 (2016).
239. Herr, A. J., Williams, L. N. & Preston, B. D. Antimutator variants of DNA polymerases. *Crit. Rev. Biochem. Mol. Biol.* **46**, 548–70 (2011).
240. Leon, D., Directed evolution of antimutator *E. coli*. PhD dissertation, University of Texas at Austin (2018).
241. Wimberly, H. *et al.* R-loops and nicks initiate DNA breakage and genome instability in non-growing *Escherichia coli*. *Nat. Commun.* **4**, 2115 (2013).
242. Shapiro, R. S., Chavez, A. & Collins, J. J. CRISPR-based genomic tools for the manipulation of genetically intractable microorganisms. *Nat. Rev. Microbiol.* **16**, 333–339 (2018).
243. Ledford, H. CRISPR, the disruptor. *Nature* **522**, 20–24 (2015).
244. Ledford, H. Alternative CRISPR system could improve genome editing. *Nature* **526**, 17–17 (2015).
245. Couce, A. *et al.* Mutator genomes decay, despite sustained fitness gains, in a long-term experiment with bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E9026–E9035 (2017).
246. van Opijnen, T., Bodi, K. L. & Camilli, A. Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nat. Methods* **6**, 767–772 (2009).